JOINT HIGHWAY RESEARCH PROJECT

FHWA/IN/JHRP-86/20

Final Report

TRAFFIC VOLUME FORECASTING METHODS FOR RURAL STATE HIGHWAYS

Jon D. Fricker
Sunil K. Saha

TRAFFIC VOLUME FORECASTING METHODS FOR RURAL STATE HIGHWAYS

FINAL REPORT

TO: H. L. Michael, Director
Joint Highway Research Project

May 27, 1987

Project: C-36-5400

FROM: Jon D. Fricker, Research Engineer
Joint Highway Research Project

File: 3-3-41

Attached is the Final Report on the HPR Part 1 Study titled, "Traffic Volume Forecasting Methods for Rural State Highways." This report presents the methods used to develop a set of traffic forecasting models for the rural highway system under the jurisdiction of the Indiana Department of Highways. The report has been prepared under the direction of Professor Jon D. Fricker.

The models fulfull the objectives of the study, in that they are well-documented, they are based on "predictor variables" that explain variations in traffic volumes, and they can be applied using only a hand calculator.

This report is forwarded to IDOH and FHWA in fulfillment of the objectives of the study.

Respectfully submitted,

Jon D. Fricker

Jon D. Fricker
Research Engineer

JDF/rrp

cc: A.G. Altschaeffl
J.M. Bell
M.E. Cantrall
W.F. Chen
W.L. Dolch
R.L. Eskew
J.D. Fricker
D.E. Hancher
R.A. Howden
M.K. Hunter
J.P. Isenbarger
J.F. McLaughlin
K.M. Mellinger
R.D. Miles
P.L. Owens
B.K. Partridge
G.T. Satterly
C.F. Scholer
K.C. Sinha
C.A. Venable
T.D. White
L.E. Wood

TRAFFIC VOLUME FORECASTING METHODS FOR RURAL STATE HIGHWAYS
FINAL REPORT

by

Jon D. Fricker
Associate Professor of Transportation Engineering

Sunil K. Saha
Graduate Research Assistant

Joint Highway Research Project

Project No.: C-36-5400

File No.: 3-3-41

Prepared as Part of an Investigation

Conducted by

Joint Highway Research Project
Engineering Experiment Station
Purdue University

in cooperation with the

Indiana Department of Highways

and the

U.S. Department of Transportation
Federal Highway Administration

Purdue University
West Lafayette, Indiana

May 27, 1987

TECHNICAL REPORT STANDARD TITLE PAGE

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.
FHWA/IN/JHRP-86-20		

4. Title and Subtitle	
TRAFFIC VOLUME FORECASTING METHODS FOR RURAL STATE HIGHWAYS	5. Report Date May 27, 1987 6. Performing Organization Code
7. Author(s) Jon D. Fricker and Sunil K. Saha	8. Performing Organization Report No. JHRP-86/20
9. Performing Organization Name and Address Joint Highway Research Project Civil Engineering Building Purdue University West Lafayette, Indiana 47907	10. Work Unit No. 11. Contract or Grant No. HPR-Part I
12. Sponsoring Agency Name and Address Indiana Department of Highways State Office Building 100 North Senate Avenue Indianapolis, Indiana 46204	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

Conducted in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

16. Abstract

Accurate forecasting of Annual Average Daily Traffic (AADT) is vital to transportation planning. The design of roads and analysis of alternative highway projects are dependent on these forecasts.

This study builds on previous efforts found in the field of rural traffic forecasting. The study combines careful statistical analysis with subjective judgment to develop models that are reliable and easy to use. This study developed two different kinds of models -- aggregate and disaggregate -- to forecast traffic volumes at rural locations in Indiana's state highway network. These models are developed using traffic data from continuous count stations in rural locations, and data for various county, state and national level demographic and economic predictor variables. Aggregate models are based on the functional classification of a highway, whereas the disaggregate models are location-specific. These models forecast future year AADT as a function of base year AADT, modified by the various predictor variables. The combination of aggregate and disaggregate models will provide reliable traffic forecasts. The number of predictor variables employed in the models was kept to a minimum. The statistical analysis also found that the predictor variables are statistically significant; no other variables will provide significant predictive power to the models. The models developed in this study provide higher R^2 values than those found in the literature, and more refined statistical techniques reinforce the choice of variables used in the models. A six-step process to obtain the future year AADT by employing both aggregate and disaggregate models is presented to assist in the model's implementation.

17. Key Words	18. Distribution Statement
traffic forecasting	No restrictions. This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

19. Security Classif. (of this report)	20. Security Classif. (of this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	288	

Form DOT F 1700.7 (8-69)

ACKNOWLEDGEMENTS

The preparation of the report was financially supported by the Federal Highway Administration and Indiana Department of Highways. The study was conducted by the Joint Highway Research Project of Purdue University in cooperation with the Indiana Department of Highways.

The principal investigator of the study was Professor Jon D. Fricker. Several individuals provided much assistance to the study. In particular, the authors are grateful to Professor Kumares C. for his encouragement, friendly support and important contributions to this report. The valuable assistance and suggestions on statistical analysis of Professor Virgil L. Anderson is gratefully acknowledge.

Special thanks are extended to John Nagle of the Indiana Department of Highways and Gary DalPorto of Federal Highway Administration for their input in this study. The authors also wish to thank all their friends who provided assistance and suggestions in course of this study.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF TABLES (continued)

LIST OF TABLES (continued)

LIST OF FIGURES

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

LIST OF FIGURES (continued)

ABSTRACT

Accurate forecasting of Annual Average Daily Traffic (AADT) is vital to transportation planning. The design of roads and analysis of alternative highway projects are dependent on these forecasts.

This study builds on previous efforts found in the field of rural traffic forecasting. The study combines careful statistical analysis with subjective judgment to develop models that are reliable and easy to use. This study developed two different kinds of models -- aggregate and disaggregate -- to forecast traffic volumes at rural locations in Indiana's state highway network. These models are developed using traffic data from continuous count stations in rural locations, and data for various county, state and national level demographic and economic predictor variables. Aggregate models are based on the functional classification of a highway, whereas the disaggregate models are location-specific. These models forecast future year AADT as a function of base year AADT, modified by the various predictor variables. The combination of aggregate and disaggregate models will

provide reliable traffic forecasts. The number of predictor variables employed in the models was kept to a minimum. The statistical analysis also found that the predictor variables are statistically significant; no other variables will provide significant predictive power to the models. The models developed in this study provide higher R^2 values than those found in the literature, and more refined statistical techniques reinforce the choice of variables used in the models. A six-step process to obtain the future year AADT by employing both aggregate and disaggregate models is presented to assist in the models' implementation.

CHAPTER 1

INTRODUCTION

1.1 Introduction

Among the most important factors in public investment decisions is the projected demand for an existing or proposed facility. The pattern of traffic growth and projected traffic volumes have been recognized as prime factors in most analyses of highway projects. Developing future traffic estimates is not an exact science, dependent as it is on so many hard-to-predict variables. The traffic growth factor has a significant effect on highway investment decisions pertaining to increasing the capacity of existing highways and the construction of new facilities, when limited funds are available. Traffic forecasting procedures must be reasonably easy and economical to carry out, be sensitive to a wide range of policy issues and alternatives, and produce information useful to decision-makers in a form that does not require extensive training to understand.

Estimates of future traffic could be arrived at by two very different methods: projections and forecasts. Projections have been used for years and are based on a historical record of the desired data item. Trend lines drawn through prior year data observations are extrapolated to the target year. In some cases these extrapolated trends are modified by the analyst based on his experience and knowledge of the route, state or region. Whereas with projections we are dealing only with the traffic data, forecasting techniques are concerned with predicting the future values of economic and other measures or indicators of person and vehicle travel. In forecasting techniques, a relationship between traffic and associated factor(s) is established.

1.2 Background

Traffic data are essential in nearly every step of the planning process. In highway investment (major maintenance, reconstruction or new construction), a reliable estimate of future traffic volume is a key element.

Traffic forecasts can be prepared with a variety approaches, depending on whether the forecast refers to an urban or rural area. In urban areas, forecasts are generally based on the four-step (trip generation model, trip distribution model, modal split model and traffic

assignment model) travel-simulation process [21,38]. In these cases travel on the road network is an output of the assignment process. Most large metropolitan areas have developed and implemented a fairly sophisticated set of computer-based travel simulation models based on the traditional four-step process. In rural areas, when assignment-based models do not exist or are not practical to apply, traffic estimates are generally made by expanding present traffic into the future based on projections of population, employment, vehicle registration, land-use data, or other parameters [21,32,38].

1.3 Past Research

Traffic forecasting in urban areas has been extensively explored and the forecasting methodologies, mainly based on sophisticated computer modeling programs, are highly advanced. On the other hand, forecasting traffic for individual rural roads, even though widely practiced, is still in its early stages. Standardized methodologies for nationwide use have not been established, and state authorities develop their own procedures to accommodate their needs. One of the reasons for the development of different procedures by different state authorities might be that, since the development of traffic projections is not an exact science, planners base

their methods on different conceptual models and thus use different procedures to reduce the uncertainty associated with their projections. Methods of traffic forecasting were advanced during the mid-sixties when statewide transportation studies were conducted by many states to fulfill the need for developing final statewide transportation plans. Traffic forecasting was a basic input for these studies.

The various state departments of highways developed their own methods to forecast rural traffic, but very few are well documented. The following sections will present some of these studies as they relate to rural traffic forecasting.

1.3.1 Traffic Growth Trends on Rural Highways

In 1958, Morf and Houska [36], in their study of the Illinois rural highway network, came to the conclusion that the four factors responsible for traffic growth patterns were (1) geographic location, (2) type and width of pavement, (3) proximity to an urban area and (4) type of service the roadway provides. They observed that growth was assumed to take the form of an S-shaped curve, as shown in Figure 1.1, with 3 stages of development -- (1) increasing growth rate (1st stage), (2) constant growth rate (2nd stage), and (3) decreasing growth rate (3rd stage).

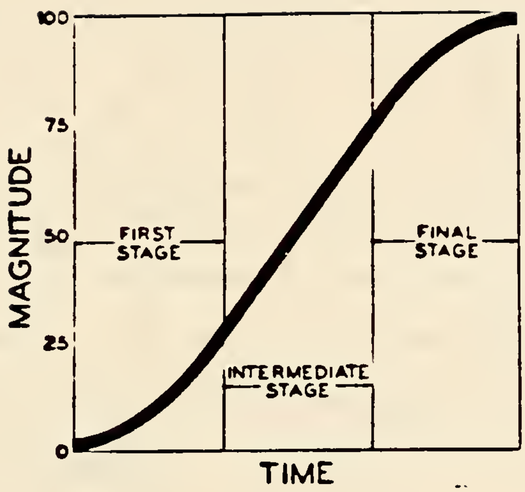


Figure 1.1: General Growth Concept

They observed that truck traffic on rural primary highways was increasing at a faster rate than passenger car traffic. Their study also indicated that population is the principal component that affects the trend, followed by persons per vehicle (or it could be expressed directly as number of vehicles) and gallons of gasoline or vehicles miles per vehicle for rural roads of Illinois.

1.3.2 Simplified Elasticity-Based Procedure

In 1982, Neveu [38] developed a set of elasticity-based models to forecast rural traffic. The models forecasted future year AADT as a function of base year Annual Average Daily Traffic (AADT), modified by various demographic factors. Neveu claimed that the type of service the roadway provides (interurban, interregional, rural to urban, urban to rural) is the only factor that had an appreciable effect on traffic growth rates. Multiple linear regression was used to identify factors that best estimated AADT and their respective elasticities. Three classes of roadway were used, as was done by Morf and Houska [36]. The background factors examined are population, number of households, automobile ownership and employment. These data are collected at town, county and state level. Neveu eliminated the income variable because of the difficulty in forecasting future values and found that the number of households is a better determinant of travel than population. Each of Neveu's models is relatively simple, with only one or two independent or predictor variables. The ultimate result of his study is a set of nomographs that give quick estimates of the growth factor, i.e., the elasticity portion of his model.

The data used for Neveu's statistical analyses were those of the year 1974 to 1978 (a total of only 5 observations for each station), in an effort to avoid any

complications from the energy crisis of the preceding years. The background data were collected for each station according to the town or county in which the station was located. The roads were classified according to the type of service they provide: (a) Interstates, (b) Principal Arterials, and (c) Minor Arterials and Major Collectors. The R^2 values 0.65, 0.77 and 0.20 for road types (a), (b) and (c), respectively, give an indication of the explanatory power of the data. For Interstates and Principal Arterials, the association of AADT with the background variables is much better than for Minor Arterials and Major Collectors. For the Minor Arterials and Major Collectors, the low R^2 indicates the poor explanatory power of the variables used. The author identifies two major problems associated with the model: (i) The difficulty in obtaining projections of the background variables and their questionable accuracy at the level they are needed, and (ii) the difficulty in deciding the applicability of the model in certain areas (i.e., whether a specific area is "rural enough" for the model).

Neveu used multiplicative constant elasticity in his model. While this specification possesses conceptual and statistical advantages, it does have an inherent weakness that should be carefully considered [28]. This weakness results from the constant elasticity structure, which implies that the effect of the growth in demand on traffic

growth always has been and will remain the same. The constant elasticity model cannot be used to forecast for more than a very limited numbers of years at a time. The result is that if the model is estimated during a period of high growth rate, future traffic will be overestimated and vice versa. Thus, when such models are used, they are recalibrated as often as practicable in order to ensure that a correction in the traffic growth rates is made and, therefore, the margin of error is limited. Models with variable elasticities are not very common in traffic forecasting. Such model structures involve more sophisticated and expensive analysis.

1.3.3 Trend Analysis-Based Procedure

The Minnesota Department of Transportation (Mn/DOT) [35] computes a route-specific growth factor from a trend analysis of the specific route. To determine the current or base year AADT, 48-hour weekday machine counts are taken and adjusted using FHWA procedures [18]. After determining base year AADT, 10-20 years (preceding to the base year) of AADT counts are taken from traffic flow maps. It has been recognized that location of the count stations on the flow map can be different from the previous year's count stations, primarily due to change of corporate limits of towns. By linear regression, a line is fitted to the data and that line is extended to the design

year. The overall growth is then the difference between design year AADT and base year AADT. Similar graphical plots of AADT against time for all (or several) major highway segments are done along the proposed project. If the growth rates are uniform, a single rate can be applied to the entire project. If not, the forecaster then must use judgment in selecting the appropriate rate for each segment based on his knowledge of the project area.

1.3.4 Disaggregate Analysis of Heavy Commercial Traffic

The New Mexico State Highway Department [2] has designed a procedure for forecasting Heavy Commercial (HC) and Average Daily Traffic (ADT) traffic on the New Mexico Interstate system and then calculating the percent HC traffic. This process, and the computer program developed from it, is called Trend-line. Trend-line identifies fourteen distinct heavy commercial truck sectors (geographical) on the New Mexico Interstate system. Separate forecasting models were developed for each sector. The disaggregate analysis (a separate analysis for each sector) provides a better traffic projection as opposed to aggregate analysis (all sectors together). Trend-line analysis includes the national, state and local socio-economic indicators that affect heavy commercial traffic on the New Mexico Interstate highways. Eight key demographic and economic indicators are identified:

1. United States Average Gasoline Cost Per Gallon.

2. United States Disposable Personal Income.

3. New Mexico Population.

4. New Mexico Residential Building Permits, Dollar Value.

5. United States Consumer Price Index.

6. United States Producer Price Index.

7. New Mexico Civilian Employment.

8. New Mexico Retail Trade.

SAS (Statistical Analysis System) multivariate analysis -- more than one dependent variable in the analysis -- was conducted using these indicators as independent variables. A series of best fit equations was developed, and percent heavy commercial of average daily traffic was forecasted for a twenty-year period.

The Trend-line sectors showed different percent heavy commercial traffic for the most recent year and led to development of separate models for the fourteen separate sectors. The state frequently uses an assumption to limit %HC to 30 percent of ADT. A regression equation that resulted in percent HC over 30 percent of ADT was defaulted back to 30 percent level.

In multivariate analysis, HC and ADT were taken as dependent variables and regressed, using socio-economic characteristics as independent variables. The socio-economic variables were identified on the national, state, county, and local level.

Once a potential indicator to estimate traffic was suggested, it was reviewed in several ways. First, it was critiqued on the basis of its theoretical applicability: How could the indicator be related to HC or ADT? The list of possible indicators was narrowed through this review. The indicators were then reviewed in several other ways: the availability of accurate information and the period of data reporting and updates.

Chi-square analysis demonstrated that ADT on the New Mexico Interstate was significantly associated with changes in state population. The standard technique of population forecasting, Cohort Analysis, was used for population forecasting and a computer program [7] was written to interface with the Trend-Line HC and ADT analysis. Cohort Analysis is the process of dividing the population into age groups, and then, each year, each age group graduates a portion into the next age group, all the babies born are added into the first age group, the different age group death rates are applied, and the net in-migration is added to forecast the next year's population. This procedure is used because different age groups have

different birth and death rates.

In the statistical analysis, linear regressions were conducted using Heavy Commercial ADT (HCADT) and ADT as dependent variables. Six years of historical data were used. The first models were multiple regression analyses of HC and ADT by year. Then multivariate analyses were done with eight independent variables. All regression analyses were conducted to find the best fit equation. All equations had an R^2 value of over 80 percent.

1.4 Scope of the Research

The purpose of this research study is to develop a method of establishing rural traffic growth factors that can be used by the Indiana Department of Highways (IDOH). The research is being carried out by the Joint Highway Research Project (JHRP) at Purdue University with the sponsorship of the Federal Highway Administration (FHWA) and IDOH.

The proposed method will be based on the background input factors for which clear relationships and usable forecasts exist and will continue to exist. Moreover, the proposed method must be reliable, well-documented and flexible. The model to be developed in this study will be simple to apply. A hand calculator will be adequate for the application of the model, making the traffic projec-

tions for any year easy to compute.

The primary focus of this study was the design and testing of a simple, fast method to forecast rural traffic volumes and step-by-step instructions on its use. This report details the development of such a procedure in order to update the method in future years. This study examines previous efforts aimed at forecasting rural traffic, describes the chosen methodology, and presents the results of the analysis. Finally, some of the limitations of the procedure are discussed, and some possible solutions to the limitations are provided.

1.5 Report Organization

This report consists of six chapters and seven appendices. Chapter 2 discusses the literature review in the light of forecasting rural traffic and the current procedures practiced by some state highway departments, as discussed in Chapter 1.

Chapter 3 addresses the problem of, and overall methodology for, constructing statistical models. Chapter 4 describes the variables in the data tables and their use in regressions.

The analysis of the data gathered in Chapter 4 is provided in Chapter 5. Statistical reliability tests are discussed in the preliminary analyses with their results.

Based on the results of preliminary analysis, two types of models (aggregate and disaggregate) are presented in Chapter 5 for different categories of highways. Chapter 5 also presents the performance of both types of models, using the data not included in the models development. Chapter 6 gives the summary and conclusions of the research as well as steps for implementation of the models developed. This chapter also provides probable problems, limitations and suggestion to overcome the problems.

The data tables for aggregate analysis developed and analyzed in Chapter 4 and 5 are presented in Appendix A. It is believed that this presentation will help in future modification of the model, if desired. Appendices B and D present the scatterplots of the dependent variable, Annual Average Daily Traffic (AADT), against the independent variables selected for aggregate and disaggregate analysis. Appendices C and E present the residual plots of the selected variables in the aggregate and disaggregate analysis. Appendix F presents four example plots of simple extrapolation. Appendix G provides the statistical test to determine the equality of two population means with an example.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

This chapter presents a review of the literature on traffic forecasting, with particular emphasis on rural traffic forecasting procedures. Some of the currently used rural traffic forecasting procedures by certain state highway departments were discussed in Section 1.3. A review of the literature reveals that limited research has been accomplished on the topic of forecasting traffic growth factors in the context of rural highways. Some ideas from this review study have been incorporated in the present study.

2.2 Transportation Demand Models

The process of relating the demand for transportation to the socioeconomic activities that generate it is known as transportation demand analysis [28]. The results of

this analysis are relationships (often in the form of models) between measures of activity and measures of transport demand. Such relationships are often referred to as transportation demand models. Although demand analysis is distinct from traffic forecasting, one can use the results of demand analysis to forecast future traffic volumes. The demand models provide a major input into the forecasting process. It should be recognized that there are limitations of demand models as forecasting tools. The strength in forecasting is not in the models or procedures used, but in the methodology applied and in the logic used to project exogenous factors. The analyst might well find it reasonable to use models of demand analysis for short-term forecasting in order to study the impacts of changes in the demand and supply environments of transportation. But as the term of forecasting becomes longer, it is unlikely that the same models will continue to be of as much relevance.

2.3 Background Factors for Rural Traffic Forecast

2.3.1 How Background Factors Affect Traffic

Memmott [32,33] showed the impact of different traffic growth rates on the estimate of future benefits from a proposed project, as well as the factors that affect traffic projection errors. These factors included

the year the projection was made, the percentage of commercial and industrial land development, and changes in highway capacity. Memmott also presented a simple model for projecting future traffic volume that is based on a multiple regression analysis of historical traffic volume data and adjustments for capacity changes and land development.

In 1980, Hartgen [20] introduced the concept of adjustment factors to base line forecasts of traffic to account for various additional concerns that had not previously been considered, or for which the previous assumptions were no longer valid. He recommended dealing with the urban and rural contexts separately. Among the aspects considered were changes in energy supply and price, auto ownership and use, households, employment and labor force, population, inflation, ridesharing, transit, and average auto fuel efficiency. He also discussed probable range of forecast errors.

2.3.2 Role of Background Factors

Covault [14] considered the impact of growth trends in population, motor vehicle registration, motor vehicle use, and motor fuel consumption on traffic growth. Hartgen [20] urged that, in nonurban areas where assignment-based modeling does not exist or may not be appropriate, estimates are generally made by extending present traffic

volumes into the future by using projections of population, number of households, cars, employment, county or town vehicle miles of travel (VMT) or other parameters. The approach taken by Hartgen [20] is to develop adjustment factors based on empirical evidence and travel elasticities those are applied to base line forecasts to obtain estimates. The factors that will influence travel are auto efficiency, gasoline price, population, energy supply cutoffs, inflation, employment, number of households, urbanization, automobile ownership and use, etc.

Salovara et al.[44] examined the impacts of background factors affecting car ownership, to prepare a forecast of traffic and the number of motor vehicles. The forecasts were compiled from three scenarios (growth, adaption and crisis) based on different international and national economic situations.

Mckay [31], in his work with Cook County and the City of Chicago, observed a close relationship between population per square mile and the amount of traffic using the highways. He found that population decreases rapidly with the increase in distance from the city of Chicago. The volume of highway traffic on each route also decreases rapidly with the increase in distance from the city. The relation between population and highway traffic indicates the necessity of considering population trends in the

formulation of a highway improvement program. The prediction of expected future traffic on the projection of the trend of motor vehicle registration is a reasonably accurate indication of future highway traffic.

Magridge [29] used forecasting car ownership as a technique to forecast traffic. The conversion of a car ownership forecast to a traffic forecast was treated as the main problem. He used two important techniques, time series analysis and cross-section analysis, to forecast car ownership. The basic assumption in a cross-section analysis, as compared with a time series analysis, is that there is a stable relationship between car ownership and income. In a subsequent article, Magridge [30] was mainly concerned with car purchases and car use. Magridge suggests that while the growth of car ownership appears likely to continue, the level of car traffic arising therefrom is much more sensitive to policy on taxation and service levels. The major determinants of car ownership are considered to be income and car prices, but not fuel prices.

2.4 Time Series Forecast of Traffic

Benjamin [6] used time series analysis to forecast future traffic. Time series analysis uses a logistic function in which model parameters are estimated by ordinary least squares. The logistic function cannot

accounts for sudden shifts in behavior or changes in transportation network, but it can provide estimates of future trends when network changes are small. Time series analysis uses land use development as the starting point to formulate the theory of traffic growth. Traffic volume is treated as a function of time and, as time passes, more land is developed and traffic increases proportionally. Land use is initially stable when the land is agriculturally zoned. As land is developed, traffic increases until all land in the zone or corridor is developed. At this point in time, traffic stabilizes. Traffic volume thereafter remains about the same, increasing or decreasing by small percentages based on variations in fuel supply, population density, driving habits and land use. The greater the land available, the greater the potential for development. Once most land is developed, there is little room for further development, so traffic growth must be slow.

The growth factor in time series analysis will be inversely proportional to the degree of land developed. The time series method of traffic forecasting is simpler and more economical than the other demand forecasting procedure and is recommended where land use is stable.

2.5 An Application of The Logistic Traffic Growth Model

Taliadoros [46] used a logistic growth model to estimate parameters to forecast traffic at ten continuous traffic count stations in Indiana. He adopted the S-shaped concept of Morf and Houska [36]. Taliadoros claimed that his procedure is simple, fast and easily calibrated with updated input data. The model he developed uses a mathematical procedure to estimate the limiting or maximum AADT and assumes that the S-curve's inflection point is a constant proportion of the limiting AADT for all stations. This study asserted that traffic data alone can provide reasonable predictions. It did not take into account any socio-economic variables and thus avoided the impact of inaccurate projections of these variables. The study does not predict temporary fluctuations in traffic growth, but only intends to project the overall growth pattern at each station.

2.6 Statewide Vehicle Counting Program

Chen [12] proposed an improved method for statewide vehicle counting program for Indiana with the help of statistical theory. The method is applicable to rural and suburban roads carrying 500 or more vehicles per day. Ritchie [43] also used a statistical approach for a better statewide traffic counting program for California. Both of these studies provide estimates of AADT that are the

basis for computing present year traffic in forecasting techniques. These estimation procedures are based on the FHWA Guide for Traffic Volume Counting Manual [18]. The data from the automatic traffic records are used to develop AADT values and monthly adjustment factors for the continuous count stations.

Drusch [16] proposed a traffic counting program to estimate AADT, which is similar to the Chen [12] study. Both of them used the FHWA method of grouping stations to convert coverage counts to AADT. Traffic counts corresponding to 24- or 48-consecutive-hours from mid Monday to mid Friday are known as coverage counts. Coverage counts are defined as single observation that, through the application of factors can be expanded to the AADT. ITE Committee 6-1 [27] looked at estimating AADT on low volume roads (less than 2000 vpd). The basis of the Chen study is Petroff and Blensly's work [40] on improving traffic count procedures by application of statistical methods. Petroff [41] earlier had developed some criteria for scheduling mechanical traffic counts, which were used later for other studies.

The expansion factor (adjustment factors) for adjusting coverage counts to AADT estimates are group mean values of monthly adjustment factors. The procedures for estimating AADT volumes used by the Indiana State Highway department, based on the FHWA "Guide for Traffic Volume

Counting Manual" [18], are:

1. A monthly adjustment factor is computed for each continuous count station for each month. It is the ratio of the AADT to the monthly average weekday traffic. The monthly average weekday traffic is computed from all the weekdays except Fridays in a month for the continuous count station.

2. The 24-hour averages of the 48-hour coverage counts are calculated. The 48-hour coverage counts are taken on weekdays, usually between noon Monday and noon Friday.

3. All the continuous count stations are grouped as per the "Guide" without considering functional grouping. The grouping steps are outline below:

   a. Using the data for the previous year arrange the monthly adjustment factors for each month in ascending order.

   b. For each month determine a set of stations such that the difference between the smallest and the largest monthly factor does not exceed the range of 0.20 in the values of the factors. For each month determine from several possible sets that set having the largest number of stations. Such a set will probably not be the same for each of

the twelve months. That is, groupings tend to vary from month to month.

c. From the twelve previously determined sets, select one set that contains those stations common to all the twelve sets. In addition to these stations, a few additional stations are assigned to the set, though they have factors that are outside of the 0.20 range in some months. Investigations have shown that special conditions can cause an abnormal change in traffic volumes for a month or two and study of the data for previous years indicated that these added stations had factors that would have placed them within the set determined from current data. A set of stations determined by such a procedure is called a group.

d. Steps b and c are repeated, considering those stations that have not been included in the first group, and a second group is selected. Steps b and c are repeated a number of times, until only those stations with extreme monthly adjustment factor values remain ungrouped. These stations are placed in a group entiled "Special Stations".

4. For each group, compute the average of the monthly

adjustment factors for each month to arrive at the group mean monthly adjustment factor. Some stations in a group, however, are not included in the computation of the mean factor for a particular month. That is, those stations having a factor outside the 0.20 range of the group for that month are not included.

5. The group mean of the monthly adjustment factors for each month is used as an adjustment factor that would be applied to 24-hour averages of 48-hour counts on weekdays.

6. The average counts, if outdated because of the 5-year cycle used in obtaining coverage counts, are updated to the current year by a traffic growth factor determined from the ATR group to which the coverage counts have been assigned.

7. The updated coverage counts are multiplied by the same year mean monthly adjustment factor of the group to which the coverage counts are assigned to obtain an estimated AADT for the roadway section where the coverage count was taken.

2.7 Comments on Forecasting Techniques

Armstrong [4,5], in his studies of forecasting, concluded that sophisticated extrapolation techniques have

had a negligible payoff for accuracy in forecasting. More sophisticated methods are generally more difficult to understand, and they cost more to develop, maintain, and implement. On the benefit side, more sophisticated methods may be expected to produce more accurate forecasts and to provide a better assessment of uncertainty. However, highly complex models may in fact reduce accuracy. While the complex models may provide better fits to historical data, this superiority does not necessarily translate into better forecasts. The danger is especially serious when limited historical data are available. He recommended simple methods and the combination of forecast techniques. The combinations may produce significant improvements in forecast reliability. The question of how many forecasts to combine is, of course, a cost/benefit issue. The weights of different forecasting method may arise another problem. Armstrong suggested starting with the least expensive method(s) and/or the most understandable method(s), and then investing in successively more expensive methods. He suggested use of methods that are as different as possible, and simply weight each forecast equally. He proposed that complexities should be avoided unless absolutely necessary. So, simple methods, which are easily understood, have been undertaken to develop traffic growth factor models in this study.

2.8 Definition of Functional Classification of Highways

The definitions of the functional classifications of rural highways [1,34] used in this study are presented below:

1. Rural Interstate: Fully controlled access facilities that are part of the interstate system. The major purpose of those highways is to provide access to and between urban areas.

2. Rural Principal Arterial: A network of routes with the following service characteristics:

   (a) Corridor movement with trip length and density suitable for substantial statewide or interstate travel.

   (b) Movements between all, or virtually all, urban areas with population over 50,000 and a large majority of those with population over 25,000.

   Thus, highways having high traffic volumes, serving the longest urban trips (one end in an urban area), and providing access to major activity centers fall in this category. In this class, service to abutting land is subordinate to the movement of traffic.

3. Rural Minor Arterial: Highways connecting with the principal arterial system and local system fall in

this category. More emphasis is placed on land access and providing service to trips of moderate length. The rural minor arterial road system, in conjunction with the rural principal arterial system, forms a network with the following service characteristics:

(a) Linkage of cities, large towns, and other traffic generators that are capable of attracting travel over long distances.

(b) Integrated interstate and intercounty service.

(c) Internal spacing consistent with population density, so that all developed areas of the state are within reasonable distances of arterial highways.

(d) Corridor movements consistent with items (a) through (c), with trip length and travel densities greater than those predominantly served by rural collector or local systems.

Minor arterials are designed to provide for relatively high travel speeds and minimum interference to through movement.

4. <u>Rural Collector</u>: Roads penetrating neighborhoods, collecting traffic from local streets, and channeling it to the arterial system. The collector system primarily provides land access. This type of road primarily serves intracounty travel and travel

distances are shorter than on arterial routes.

5. Rural Local Road: Roads providing direct access to abutting land. Through traffic usually does not use this type of road. These local roads serve travel over relatively short distances.

2.9 Chapter Summary

The objective of this chapter is to provide a brief review of the literature pertaining to rural traffic forecasting. Definitions of the functional classifications of rural highways have been provided to aid in classifying a highway for which a traffic growth factor is desired. Procedures to estimate AADT from short-term traffic count have been introduced. The commonly-cited background factors for rural traffic forecasting have been identified in this chapter. Some of those factors, for which data are available, are used in this research.

CHAPTER 3

PROBLEM STATEMENT AND METHODOLOGY

The forecasting of traffic on rural highways has not been a major focus of transportation research. Most of the critical issues of this area have already been mentioned in Section 1.3 and in Chapter 2 (Literature Review). In this research, an effort is made to develop models to predict future traffic on rural highways in Indiana.

The current practice at the Indiana Department of Highways (IDOH) to forecast future traffic on state highways is based on a pair of 20-year growth factors for each of Indiana's 92 counties. One growth factor in a county applies to its rural highways, the other to its urban sections. Recognizing that the current set of traffic growth factors are outdated, overly simplistic, and lacking the documentation necessary to update them, the proposed method will provide a means of predicting future traffic volumes that is reliable, well-documented,

flexible, and based on input factors for which clear relationships and usable forecasts will continue to exist.

A clear distinction should be made about the nature of traffic forecasting methodologies. They are divided into two separate groups: (1) Those that address the forecasting problem as a network analysis, based on traditional four-step process that requires enormous amounts of data and sophisticated computer resources, while not guaranteeing forecasts that are appreciably superior to less detailed methods, and (2) the simple, easy-to-use forecasts on a road-to-road basis that fulfill the particular needs of the local highway departments.

The proposed method seeks a suitable "middle ground" -- a method that provides a reliable forecast with modest data and computational requirements. The models developed should be relatively simple to use and could be updated without difficulty. This study will meet the continuous needs of IDOH for a reliable method of estimating future traffic on individual routes as an aid to the planning process and in implementing the Highway Improvement Program.

The study by Morf and Houska [36] leads to the conclusion that the characteristic "type of service" has a remarkable effect on traffic growth rates. Highways with the greatest percentage of interurban or interregional

service generally had the largest increases in travel. Roads that serve largely urban-to-rural or rural-to-urban travel had the smallest increases. The results of the Morf and Houska study suggest different traffic forecasting models for different functional classes of highway. The functional classification of highway are interstates (representing interurban and interregional service), principal arterials (representing rural-to-urban service), and minor arterials and major collectors (representing rural-to-rural service). By using functional class as the determinant, the four road types were rural interstate, rural principal arterial, rural minor arterial, and rural major collector. Statistical analyses in Chapter 5 suggest a different model for each of the four highway categories and/or a separate model for each station, as opposed to one common model for all highway categories.

A variety of forecasting models were examined. The simplest one was AADT (Annual Average Daily Traffic) being directly proportional to the background factors of Table 3.1, such as population or number of households. Table 3.1 presents a summary of background factors used in developing the models. State level data were used only in case of interstate and principal arterial highways. However, it was felt that the explanatory power of such a model would be too low to provide reasonably accurate forecasts. Such a procedure also carries with it a problem

Table 3.1

Background Factors

Level	Factors
A. County	1. Population 2. Households 3. Vehicle Registrations 4. Employment 5. Income
B. State	1. Population 2. Households 3. Vehicle Registrations 4. Employment 5. Income
C. National	1. Gasoline Price 2. Consumer Price Index 3. Gross National Product 4. Income
D. Other	Year

inherent in all regression models: the problem of forecasting outside the range of predictor variables in which it was calibrated.

An elasticity-based model [38] was finally selected and used to relate future year AADT to present year AADT by means of a number of background factors. The general form of the model is as follows:

$$AADT_f = AADT_p \left| 1.0 + \sum_{j=1}^{n} e_j (x_{j,f} - x_{j,p})/x_{j,p} \right| \qquad (3.1)$$

or, upon rearrangement,

$$\frac{AADT_f - AADT_p}{AADT_p} = \sum_{j=1}^{n} e_j \left| \frac{x_{j,f} - x_{j,p}}{x_{j,p}} \right| \qquad (3.2)$$

where,

$AADT_f$ = AADT in future year,

$AADT_p$ = AADT in present year,

$x_{j,f}$ = value of variable x_j in the future year,

$x_{j,p}$ = value of variable x_j in the present year,

e_j = elasticity of AADT with respect to x_j,

n = number of associated variables.

The elasticity-based model was selected for several reasons. The most important reason was that it was believed that the range of volumes over which the model would be applied would be much greater than that used in developing the model, making a simple linear regression

model that relates AADT to the background factors directly inappropriate. Second, the use of present year AADT to estimate future year AADT (as a sort of pivot point) would reduce the problem of nonresident travel. Also, the elasticity portion of the model calculates a growth factor directly (See right hand side of equation 3.2).

The AADT values were obtained from the Highway Department's continuous count program. Only those stations classified as rural in nature were selected for use in the study. This yielded a total of 23 stations throughout the state for the four categories of highways. Those stations are shown in Figure 3.1. Based on the county in which the automatic traffic count station is located, the various background factors (see Table 3.1) were collected.

The elasticities and the appropriate background factors are derived from a linear equation that relates AADT to a variety of the factors in Table 3.1. It can be shown mathematically that, given an equation of the form:

$$Y_i = a + \sum_{j=1}^{n} a_j X_{ij} \qquad (3.3)$$

where,

Y_i = value of dependent variable at ith observation; $i = 1,\ldots\ldots,n_1$,

X_{ij} = value of jth independent variable at ith observation; $j = 1,\ldots\ldots,n$,

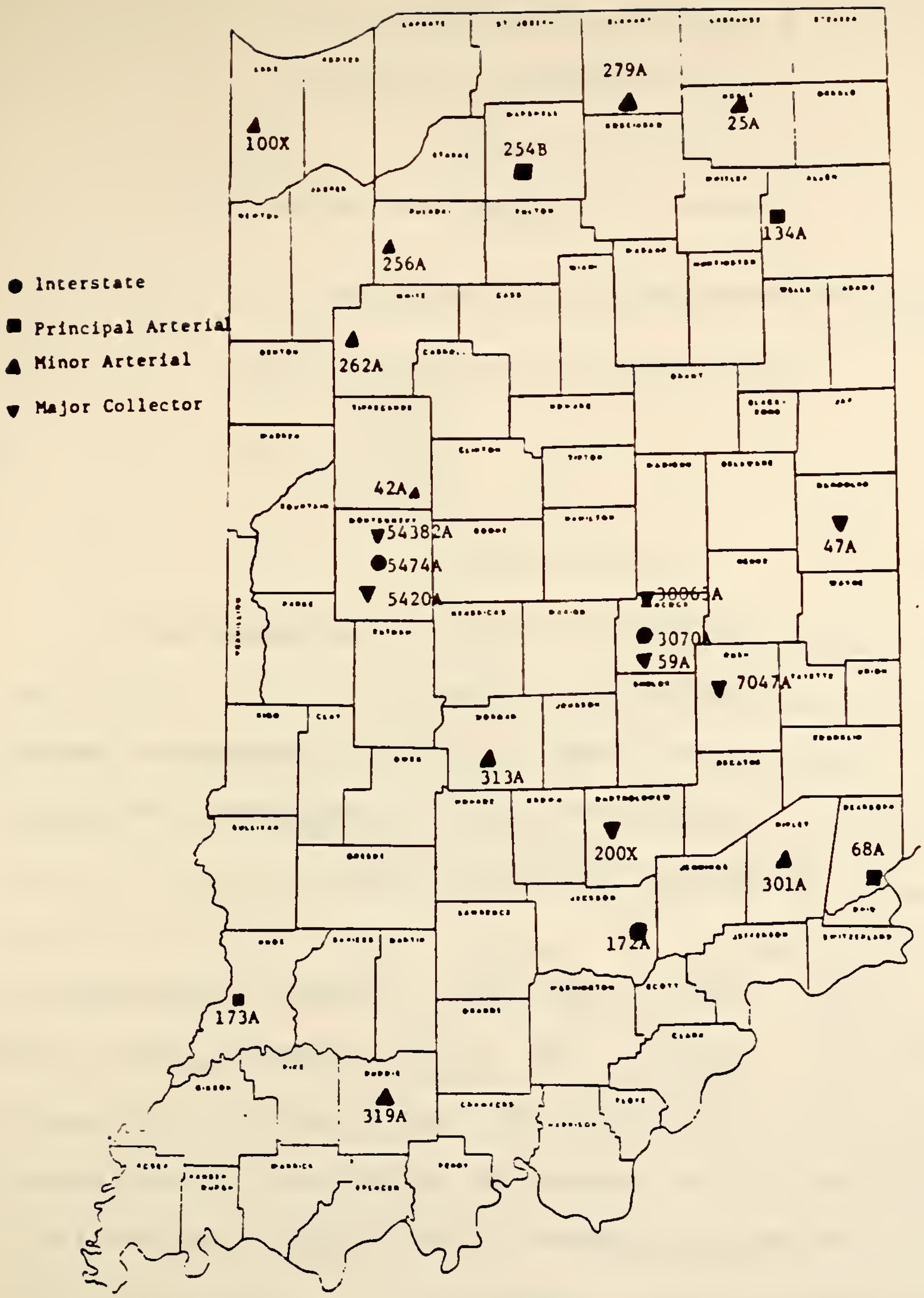


Figure 3.1: Rural Continuous Count Stations, State of Indiana

a = constant term,

a_j = regression coefficient for jth independent variable,

n_i = observation number,

n = number of independent variable.

Elasticity measures can be estimated by:

$$e_j = a_j \left| \frac{\overline{X}_{ij}}{\overline{Y}_i} \right| \tag{3.4}$$

where,

e_j = elasticity of AADT with respect to independent variable x_j,

$\overline{X}_{ij}$ = overall mean of the jth independent variable,

$\overline{Y}_i$ = overall mean value of dependent variable,

a_j as defined below equation 3.3.

Thus, using multiple linear regression, the background factors that best estimate AADT and their respective elasticities can be derived. The data for estimation of the background factors and elasticities came from a variety of sources. Details regarding the data are presented in the next chapter.

CHAPTER 4

DATA COLLECTION AND DATA TABLES

4.1 Introduction

In this chapter, a number of variables that have been identified in Table 3.1 will be discussed along with the traffic data for which forecasts are desired. The data tables for different highway categories, identified in earlier chapters, will also be discussed. These data tables are the input medium for statistical analysis. Some of the earlier attempts, which were dropped later on due to some difficulties, are described briefly in this chapter. The main objective of this chapter is to describe the variables and the evolution of the data tables used in the analysis. The sources of the data and their conversion, where needed, are discussed in detail. These data tables could be modified when new count stations and/or new census reports become available, in order to calibrate and modify the developed models to predict future traffic.

The variables examined by regression analysis are shown in Table 4.1.

4.2 Description of Variables

Y, Annual Average Daily Traffic (AADT)

AADT is the average 24-hour traffic volume for a given year, for both directions of travel, unless otherwise specified. This is the only response variable which needs to be predicted in future years. The State of Indiana has altogether 23 rural continuous traffic count stations (identified in Figure 3.1 of Chapter 3) to measure AADT on different functional classes of highway [23]. For this study, each station has been assigned to one of the four categories of highway identified in Chapter 2. The resulting classification is shown in Table 4.2.

In the early stages of this study, data tables were based on traffic data from the 1950's, 1960's and 1970's. In those cases, the data for every fifth year were taken. The aim in these early stages was to use only basic (easily acquired) census data as the independent variables. However, the literature [5,15,37,38] suggests the use of only recent data to develop model(s) for forecasting.

Table 4.1

Variables for Regression Analysis

Symbol	Description of the Variable	Type of Variable
Y	Annual Average Daily Traffic (AADT)	---------
X_1	County Vehicle Registrations	Demographic
X_2	US Gasoline Price in cents per gallon, 1972 $	Economic
X_3	Year	---------
X_4	County Population	Demographic
X_5	County Households	Demographic
X_6	County Employment	Economic
X_7	State Vehicle Registrations	Demographic
X_8	State Population	Demographic
X_9	State Households	Demographic
X_{10}	State Employment	Economic
X_{11}	Consumer Price Index (CPI) - US	Economic
X_{12}	Gross National Product (GNP), in billions of 1972 dollars	Economic
X_{13}	Per Capita Disposable Personal Income (national), in 1972 $	Economic

Table 4.2

IDOH (*) Rural Continuous Stations: Location and Highway Category

Highway Category	Count Station	Highway Name	Station Location (County)
1. Rural Interstate	172A	I-65	Jackson
	3070A	I-70	Hancock
	5474A	I-74	Montgomery
2. Rural Principal Arterial	68A	US 50	Dearborn
	134A	US 30	Allen
	173A	US 41	Knox
	254B	US 31	Marshall
3. Rural Minor Arterial	25A	SR 9	Noble
	273A	US 6	Elkhart
	301A	US 421	Ripley
	313A	SR 67	Morgan
	319A	SR 56	Dubois
	42A	US 52	Tippecanoe
	100X	US 41	Lake
	256A	US 421	Pulaski
	262A	US 24	White
4. Rural Major Collector	47A	SR 1	Randolph
	7047A	CR 68 (900N)	Rush
	30063A	CR 63 (600E)	Hancock
	54382A	CR 382 (400W)	Montgomery
	59A	US 40	Hancock
	200X	US 31	Bartholomew
	5420A	US 136	Montgomery

(*) IDOH — Indiana Division of Highway

An attempt to increase the number of cases or observations for each category of highway was made. The use of annual data, as opposed to every fifth year data, helped to increase the number of cases. To expand the number of observations, the following procedure was used. Traffic flow maps [25] were closely examined with the help of the 1985 functional classification system map of Indiana. Several problems resulted from the use of these traffic flow maps. First, the counts indicated on different traffic flow maps were for highway segments whose end points would vary with each edition of the map. Second, the traffic estimates on the traffic flow map are dependent on some adjustment factors derived from traffic counts at continuous count stations. They are not pure volumes taken under constant conditions but are themselves estimates. Finally, the traffic data on the flow maps need interpolation to determine the traffic for years other those in which the such flow map data were assembled. Consequently, the idea of using traffic data from the traffic flow maps was dropped in favor of interpolated values. At that point, the development of a prototype model took precedence over precise values for each year at continuous count stations. The traffic data (AADT) [23] used in this study are being taken from the 23 rural continuous count atation for the years 1970 to 1984.

AADT data from 1970 to 1982 were used to develop the data tables, providing as many as thirteen AADT observations per count station. The traffic data for 1983 and 1984 were not used in developing the model(s), but were kept aside to test the model(s) to be developed. Column 1 of the data tables in Appendix A contains AADT (Y).

X_1, <u>County Vehicle Registrations</u> and
X_7, <u>State Vehicle Registrations</u>

The total number of vehicle registrations in the county where a count station is located (X_1) and that for the whole state of Indiana (X_7) is published each year by the Indiana Bureau of Motor Vehicles [22]. These data are reliable in the sense that they are not estimates, but are counts made at motor vehicle registration offices throughout the state. These variables are proposed to explain AADT (Y) on the assumption that AADT in a particular year at a given place is closely related to the number of vehicles registered then and there. The prediction of expected future traffic based on the projection of the trend of motor vehicle registrations is a reasonably accurate indication of future highway traffic. The value of variable X_1 for each year, 1970 to 1982, was used in the data tables for all categories of highways. The variable X_7 was used only for Interstates and principal arterials, because it was believed that state level data influence those highways that run across

the state. Column 3 of the data tables in Appendix A represents X_1, and column 9 of Tables A1 and A2 in Appendix A represents X_7 for the years 1970 to 1982.

X_2, US Gasoline Price in cents per gallon, 1972 dollars

The variable X_2 was used for all categories of highway, on the assumption that the price of gasoline at the state and county level parallels the national level retail price. For use in the data tables, the prices (see Table 4.3) were converted to 1972 dollars by applying the consumer price index (CPI) for transportation [8,17] in equation 4.1.

$$P_{1972} = \left| \frac{CPI_{1972}}{CPI_{19XX}} \right| P_{19XX} \qquad (4.1)$$

where,

P_{1972} = US retail motor gasoline price in cents per gallon, in 1972 $, for the year 19XX,

P_{19XX} = US retail motor gasoline price in cents per gallon, in current $, for the year 19XX,

CPI_{1972} = Consumer Price Index for transportation for the year 1972 (CPI_{1967} = 100),

CPI_{19XX} = Consumer Price Index for transportation for the year 19XX.

Table 4.3

US Gasoline Prices [45]

U.S. gasoline prices

(Retail motor gasoline — cents per gallon)

Year	Before tax	Taxes	Total	Year	Before tax	Taxes	Total
1983	104.86	14.65	119.51	1973	26.88	11.94	38.82
R 1982	111.29	14.23	125.52	1972	24.46	11.67	36.13
R 1981	118.69	12.99	131.68	1971	25.24	11.24	36.48
1980	110.95	12.85	123.80	1970	24.55	11.14	35.69
1979	72.95	12.75	85.70	1969	23.85	10.99	34.84
1978	49.98	12.82	62.80	1968	22.93	10.78	33.71
1977	49.83	12.37	62.20	1967	22.55	10.60	33.16
1976	46.97	12.03	59.00	1966	21.57	10.51	32.08
1975	44.93	11.77	56.70	1965	20.71	10.46	31.17
1974	41.20	12.00	53.20	1964	19.98	10.37	30.35

R — Revised

Source: Independent Petroleum Association of America

The converted US gasoline prices in 1972 dollars (X_2) are shown under column 4 in the data tables of Appendix A. This variable is adopted on the assumption that AADT is inversely proportional to the price of gasoline.

X_3, Year

The variable X_3 represents the year in which all the variables, both dependent and independent, apply for a particular observation, i.e., the row in the data tables of Appendix A. This is simply the year, shown in fifth column in data tables of Appendix A as 1970, 1971,...,

1982. This variable was introduced to reflect the time effect on AADT (Y) and the X-variables, to study the residual patterns against time. As a general trend, AADT increases as time passes. It is assumed that using X_3 as a variable will lead to high statistical correlation with the other predictor variables (X's) and, in that event, X_3 could be dropped from the models.

X_4, <u>County Population</u> and
X_8, <u>State Population</u>

Decennial Bureau of Census records [9,11] on population are used for the state and county values. These variables are taken as predictor variables on the assumption that the response variable Y (AADT) in a particular year at a place is dependent on the number of people living there. The variable X_8 was used only for Interstates and principal arterials, because it was believed that state level data influence those highways that run across the state. Intercensus estimates of X_4 and X_8 from the Indiana School of Business [26] were used in the data tables for years other than census years. The Indiana Business School also projects the population for every fifth year in the intermediate future. Its projections are made at the county level, based on the fertility, mortality, and net migration experiences of the county populations. The state forecasts are the results of the sum of the forecasts of the 92 individual counties. The projec-

tions are based on past trends and patterns, but also involve judgments, because simple historical extrapolation is not always reliable. Column 6 of the data tables in Appendix A represents X_4, and column 10 of Tables A1 and A2 in Appendix A represents X_8 for the years 1970 to 1982.

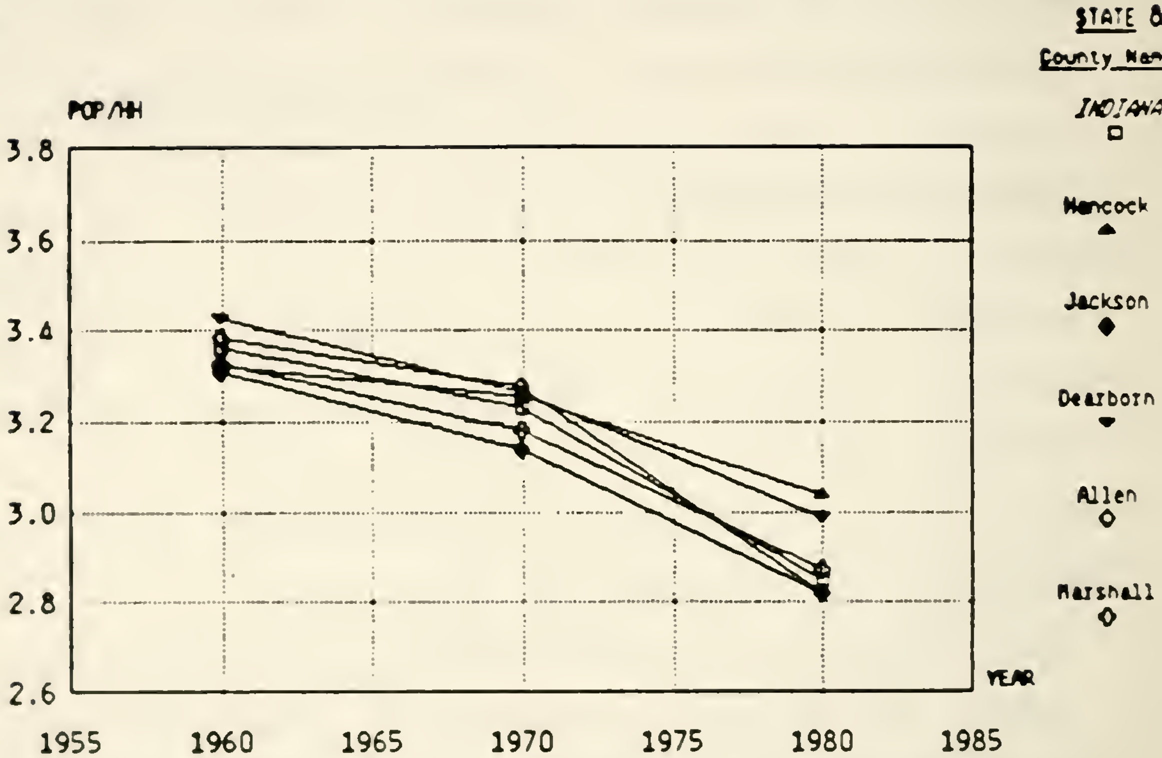
X_5, <u>County Households</u>, and
X_9, <u>State Households</u>

A household includes all persons who occupy a housing unit. A housing unit is a house, an apartment, a group of rooms, or a single room occupied as separate living quarters or, if vacant, intended for occupancy as separate living quarters. Data for total households include all occupied housing units. The number of occupied housing units is the same as the number of households. The housing statistics presented here for the years 1970 and 1980 are based on the results of the 1970 and 1980 Census of Population and Housing, conducted by the Bureau of Census as of April 1, 1970 and 1980 [9]. Some of the data collected by the Bureau of Census were collected on a 100 percent, or complete-count, housing inventory, while other data were obtained from sample estimates. The samples were of 5 percent, 15 percent, and 20 percent, depending on the subject covered. The sample data have been "weighted" or "inflated" to reflect the entire population or universe. Exact agreement, therefore, is not to be expected between data based on samples and data resulting

from complete counts.

The total number of households in a county in a particular year is X_5 for that year. X_9 is the statewide value. It was found in Neveu's study [38] that number of households is a better estimate of AADT than population. The predictor variables (X_5 and X_9) are chosen on the assumption that the response variable Y (AADT) will be adequately explained by using them in models. The variable X_9 was used only for Interstates and principal arterials, because it was believed that state level data influence highways that run across the state. The Bureau of Census [9,11] gives the values of X_5 and X_9 for each census year, while and the Indiana Business School makes projections of households for each county. The state-level projection is then simply the sum of the 92 county values. The estimates of intercensus households between 1970 and 1984 were accomplished by the procedure described below.

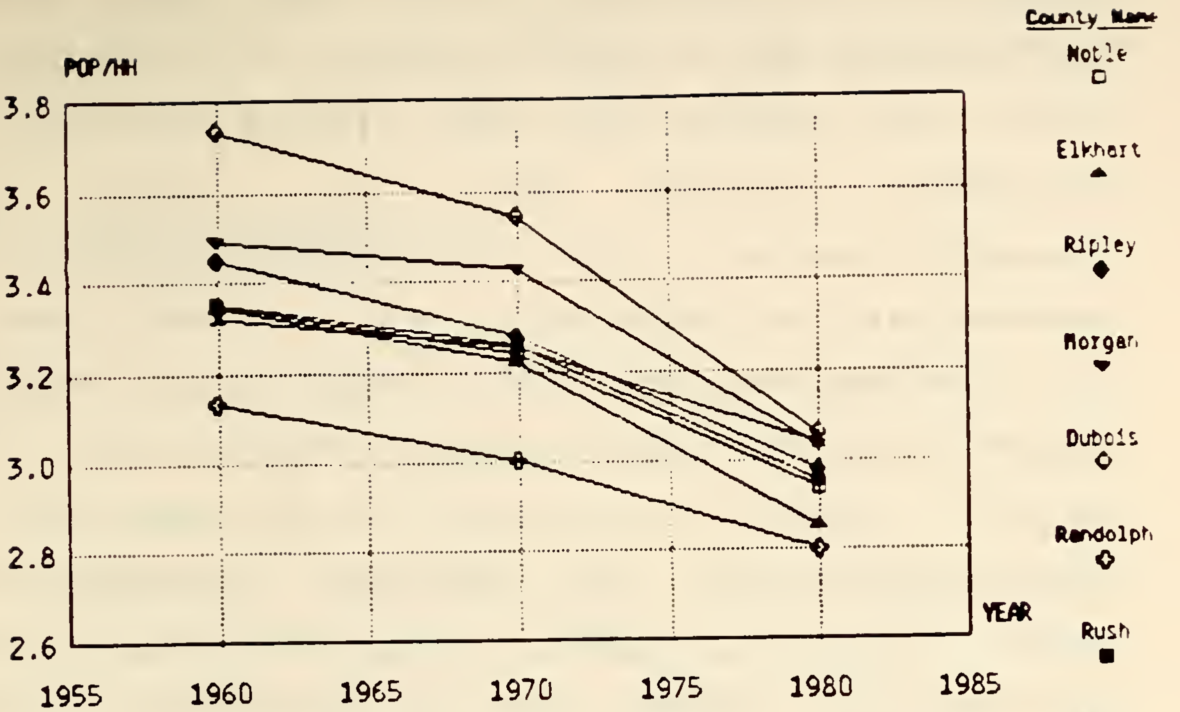
Figure 4.1 presents the ratio of population and households for the past three census years: 1960, 1970, 1980. State and the counties with Automatic Traffic Record (ATR) stations are shown on the plots of Figure 4.1. The figure indicates that the slope for 1970 to 1980 is greater than that for 1960 to 1970. Although not shown in Figure 4.1, the slope in some counties was positive in the decade 1950 to 1960. It is assumed from the nature of



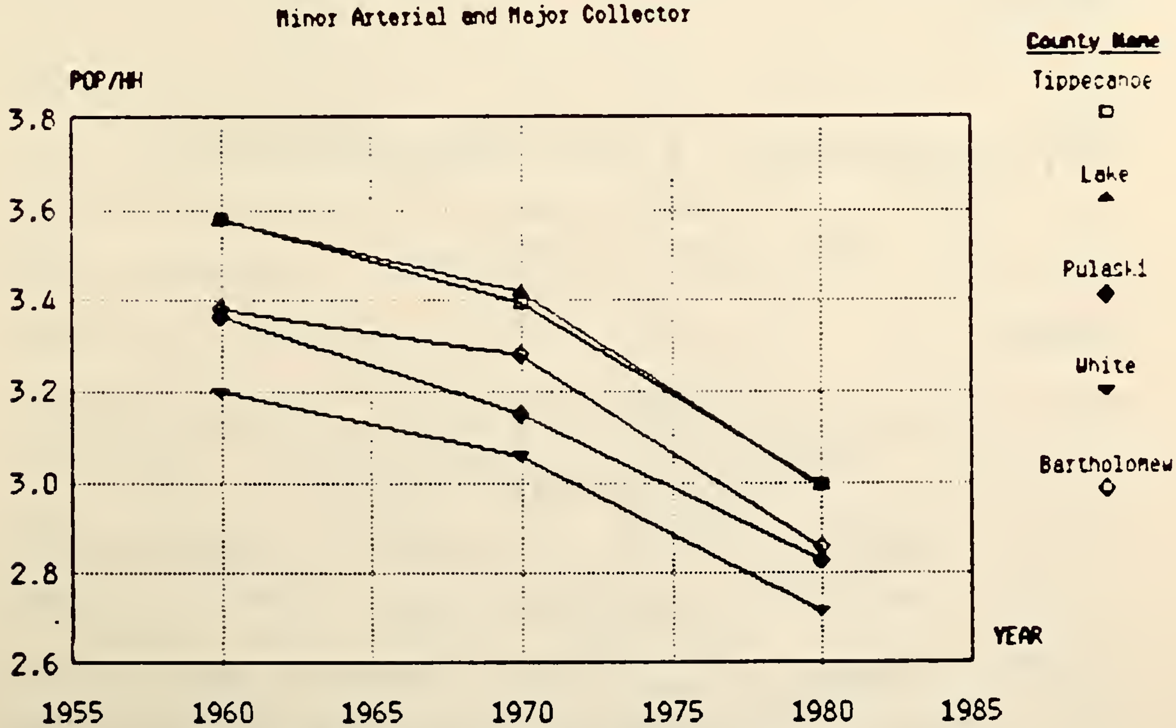
(a) State and Counties with ATR stations for Rural Interstate and Principal Arterial

Figure 4.1: Ratio of Population to Households for census year 1960, 1970 and 1980

Note: ATR = Automatic Traffic Record
POP/HH = Population per Household



(b) Counties with ATR stations for Rural Minor Arterial and Major Collector



(c) Counties with ATR stations for Interrupted Rural Minor Arterial and Major Collector

Figure 4.1 (continued)

the curve that the average household size (total population/total households) will not change significantly for the decade 1980 to 1990 with respect to its earlier decade. The general trend will continue to be one of decreasing household size. The slopes of population/household between 1970 and 1980 are less than 0.04 per year (see Figure 4.1). With these mild slopes, it is assumed that the average household size is changing uniformly between the census years and that the same rate could be expected for the next 3 or 4 years after a census. Based on the above assumptions, households at years 1971 to 1979 and 1981 to 1984 are computed by using equation 4.2 for the whole state and for counties with ATR stations.

$$HH_{19XX} = \left| POP/HH_{1970} + \left| \frac{POP/HH_{1980} - POP/HH_{1970}}{10} \right| (19XX - 1970) \right|^{-1} POP_{19XX} \qquad (4.2)$$

where,

POP/HH_{1970} = Ratio of population to households in the year 1970,

POP/HH_{1980} = Ratio of population to households in the year 1980,

POP_{19XX} = Population in the year 19XX ($1971 < 19XX < 1979$ and 1981,...,1984),

HH_{19XX} = Households in the year 19XX.

The column 7 of the data tables in Appendix A represents X_5, and column 11 of Tables A1 and A2 in Appendix A represents X_9 for the years 1970 to 1982.

X_6, <u>County Employment</u> and
X_{10}, <u>State Employment</u>

Employment data [11,13,24] is an economic variable. The County Employment Patterns [24] are released each summer and provide "covered employment" data for each month, each county, and each employment category for the previous calendar year. County Employment Patterns published prior to 1983 do not provide state employment figures. According to the 1983 edition, total "covered employment" consists of 1. Mining, 2. Construction, 3. Manufacturing: (a) Food, (b) Textiles, (c) Lumber, Wood Processing, (d) Furniture, (e) Paper, (f) Printing, (g) Chemicals, (h) Petroleum Products, (i) Rubber, Plastics (j) Leather, (k) Stone-Clay-Glass, (l) Primary Metals, (m) Fabricated Materials, (n) Non-electric Machinery, (o) Electric Machinery, (p) Transportation Equipment, (q) Instruments and (r) Misc. Manufacturing, 4. Transportation, Communication, Public Utilities, 5. Wholesale Trade, 6. Retail Trade, 7. Finance, 8. Agriculture & Services, and 9. Government. According to the 1976 edition, covered employment represents about 85 percent of nonagricultural wage and salary employment and 78 percent of all employment. Major exceptions to coverage of wage and salary

employment are in railroads, small nonprofit institutions, churches, private households, and most government units. State hospitals, schools of higher education, and local government utilities are covered. In addition to these exceptions, self-employed workers (both farm and non-farm) are excluded from coverage.

"County Business Patterns - Indiana" [10], a publication of the US Bureau of Census, furnishes employment data for each year for the week including March 12, and provides such data for the county and state levels. This summary of employment excludes government employees, railroad employees, self-employed persons, etc. This publication also provides Federal Civilian Employment for the mid-March pay period by county and state. The "City and County Data Book" [9] is another publication of the Bureau of Census that presents employment data by county and state in every tenth year. The employment figures in the "City and County Data Book" are prepared from household surveys, where workers are counted according to their place of residence; whereas for "County Business Patterns", they are counted according to their place of work. There are various reasons for differences in the two series of data: differences in the reporting systems they use; differences in the time period to which the reports refer; sampling variations in the figures based on the sample survey and differences in industrial classification

resulting from the fact that the survey information is obtained from respondents in workers' households, whereas the County Business Patterns industrial classification is based upon information either from the employer or administrative sources.

There exists little difference between the numbers in "County Business Patterns" and in "County Employment Patterns". These differences are mainly due to the reporting systems and the periods to which the reports refer. The average yearly employment numbers from the "County Employment Patterns" [24] were taken as variable X_6, County Employment. The state employment (X_{10}) are taken by summing two tables, 1E and Appendix, of County Business Patterns [10]. Table 1E [10] gives the number of employees for the week including March 12 and excludes government employees, railroad employees, self-employed persons, etc. The Appendix table [10] gives the number of federal civilian employees in the mid-March pay period. The variable X_6, column 8 of the data tables in Appendix A, was used in all types of highways, but the variable X_{10}, column 12 of Tables A1 and A2 in Appendix A, was used only in case of Interstates and Principal arterials.

X_{11}, Consumer Price Index (CPI) - US

This is an economic indicator at the national level. The data for the consumer price index [8,17] are for the

US city average. The CPI data are used in case of Rural Interstates and Rural Principal Arterials, column 13 of Tables A1 and A2 in Appendix A. The CPI value at 1967 has been taken as 100 and all other years' data have been expressed with respect to this base year. The CPI values represent economic comparison at different years and it is believed that AADT at different years are correlated with this economic indicator.

X_{12}, Gross National Product (GNP), in billions of 1972 dollars

These data are measure of the value of goods and services in the nation. It is believed that traffic (especially truck traffic) on Interstates and principal arterials will be explained by GNP, X_{12}. This variable is presented in column 14 of Tables A1 and A2 in Appendix A. The data for GNP in billions of 1972 dollars were obtained from a monthly publication entitled "Economic Indicators" [13] and are available for each year.

X_{13}, Per Capita Disposable Personal Income (national), in 1972 dollars

This is also a national level economic indicator and is used only in the case of Rural Interstates and Rural Principal arterials, presented in column 15 of Tables A1 in Appendix A. It is believed that this national level income influences the traffic at national highways. This variable was used earlier in New Mexico's [2] and Neveu's

[38] study. Disposable personal income represents the income after personal taxes and nontax expenditures. The data for each year of per capita disposable personal income were presented in "Economic Indicators" [13] both in current dollars and in 1972 dollars.

The City and County Data Book [9] publishes the per capita personal income and median family income at the state and county level for the year before the census years. An estimate is required for other years. An attempt to estimate incomes by graphical interpolation was found to be unreliable. Moreover, future values for either of these income variables are difficult to forecast, especially in an economy that is subject to rapid changes. Based on the stated criterion of using independent variables that are easily available and simple to forecast, the present data tables for 1970 to 1982 exclude any income variables at the state and county levels from consideration. The national level data are readily available from "Economic Indicators" [13], where the data are presented for each year. The future value of this national level income in 1972 dollars can be reasonably estimated by extrapolating the graphical plot -- income vs. year -- of the data from "Economic Indicators" [13].

4.3 The Data Tables

The four data tables for aggregate analysis, one for each category of highway identified in Table 4.2, are presented in Appendix A as Tables A1 through A4. Those stations in a functional category whose data were clearly well out of the range of values for most of the stations in its category were not used in the development of an aggregate model. Instead, these stations were "saved" to test the ability of an aggregate model to "predict" their AADT values. The variables X_7 to X_{13} were used as candidate background factors only in the case of Rural Interstates and Rural Principal Arterials. The variables X_1 to X_6 were candidates in all highway categories. Each row or case of Appendix A corresponds to the year given under column 5. The tables of Appendix A are labeled in rows to identify a row or observation that corresponds to a Automatic Traffic Record (ATR) count station.

The data tables present all possible cases or observations for ATR count stations in rural Indiana between 1970 and 1982. The resulting number of cases were 26 Rural Interstates, 39 principal arterials, 52 minor arterials and 37 major collectors respectively.

The data tables will be analyzed in two ways -- by using disaggregate and aggregate techniques. In disaggregate analysis, each station (including those dropped from

the aggregate data tables) will be analyzed separately. Station- or location-specific models for highways having similar characteristics will be developed. In aggregate analysis, stations under a given category of highway will be analyzed as a group, and a model applicable to any highway classifiable within a certain group will be proposed. The value of each approach for each highway type will be assessed through some trial forecasts in Chapter 5.

4.4 Chapter Summary

The central idea of this chapter is to describe the variables used in model development. The variables identified in Chapter 3 have been discussed and the sources of their numerical values are given. Explanations behind the uses of all the predictor variables (or independent variables, X´s) are given. The methods by which certain data are estimated or converted to a form compatible with the proposed model are presented. The reasons behind dropping some variables from consideration have also been briefly discussed. Some of the earlier attempts at data acquisition are also presented. This chapter is a guide to the data tables appearing in Appendix A.

CHAPTER 5

STATISTICAL ANALYSIS AND MODEL DEVELOPMENT

The statistical analyses of the variables identified in Chapters 3 and 4 are described in this chapter. First, models to predict future traffic, based on the data tables of Appendix A, are developed. The performance of these models are then tested by trying to predict the observations that were not included in the development of the model.

5.1 Introduction to Statistical Analysis

As was mentioned in Chapter 3, in order to develop a reasonable causal relationship, a regression procedure that fits a least square estimator of AADT to the background variables is the basis for the development of the model. The regression approach was selected because: (1) the SPSS package permits computation of the elasticity of the dependent variable (in this case, AADT) with

respect to the independent variables, (2) it provides an estimate of the function regressed (here, AADT) that could also be used for prediction purposes in the future, and (3) regression allows, by means of linear tests associated with it, testing the significance of the effects of different variables (X's) in the equation.

5.2 Preliminary Analysis

The preliminary statistical analyses were done to identify any possible relationship between dependent (Y) and independent variables (X's) through scattergrams and to check the normality and homogeneity of variance assumption in the regression approach.

5.2.1 Homogeneity of Variance

The Statistical Package for the Social Sciences (SPSS) one-way program [39] was used to identify the homogeneity of variances of AADT between the stations in a category of highway for an equal number of observations in each station or group. The homogeneity of variance of the AADT was checked using the Cochran and Bartlett-Box tests [3] by treating the Y's for each station as a group for an equal number of observations in each station or group. The Burr-Foster Q-test [3] was performed to check the homogeneity of variance of Y's at different stations for a highway category with an unequal number of observations

among stations. The q statistic for the Burr-Foster Q-test for unequal sample sizes was calculated from equation (5.1).

$$q = \frac{\bar{\nu} \sum_{i=1}^{P} (\nu_i S_i^4)}{\left| \sum_{i=1}^{P} \nu_i S_i^2 \right|^2} \tag{5.1}$$

where,

ν_i = Degree of freedom for ith station or group ($\nu_i = n_i - 1$ and $i = 1,\ldots\ldots,P$),

n_i = Number of observation for ith station,

P = Number of stations or groups,

S_i^2 = Sample variance for the ith station or group,

$\bar{\nu}$ = Arithmetic average of degrees of freedom.

No one has come up with a β-level for homogeneity tests that will indicate when the experimenter or researcher should become concerned about making a transformation. But a set of working rules that seem to be effective for the practitioner has been advanced [3]:

1. If the homogeneity test is accepted at the $\beta = 0.01$ level, tranaformation is not needed.

2. If the test is rejected at the $\beta = 0.001$ level, transformation is needed.

3. If the result of the homogeneity test is somewhere between $\beta = 0.01$ and 0.001 and if there is a practical reason to transform, then the usual transformation can be done; otherwise, it is recommended not to perform the transformation.

4. If the investigator has no theoretical knowledge of his variable, it is recommended to use $\beta = 0.001$ when the distribution of Y seems to have excessively long tails.

The test results for homogeneity of variance in Table 5.1 show that the highway categories 1, 2 and 3 satisfy the homogeneity of variance condition by both the Cochran and Bartlett-Box tests. In case of Rural Interstates, the β-level for the Cochran test was found to be greater than 0.05. Thus, the homogeneity of variance for Rural Interstates is satisfied for Cochran β-level of 0.05. A β-level greater than 0.01 for the Bartlett-Box test satisfied homogeneity of variance for Rural Interstates at a β-level of 0.01. But the β-level for the Bartlett-Box test for Rural Minor Arterial is 0.001. Based on the regression analysis, a linear relationship between Y and the X's is feasible. There is no apparent practical or theoretical reason to transform Y. The distribution of Y's at some stations is sparse, as indicated in the data tables. Considering all these factors,

Table 5.1

Results of the Tests for Homogeneity of Variance

A. Bartlett-Box and Cochran Test (Equal Sample Size)					Homogenity of Variance
Highway Category (No. of station or group)	Cochran c [β-level]	Bartlett-Box f [β-level]	Remarks on β-level		
			Cochran	Bartlett-Box	
1. Rural Interstate (2)	0.6049 [0.471]	0.519 [0.471]	$\beta > 0.05$	$\beta >> 0.01$	Checked
2. Rural Principal Arterial (3)	0.5686 [0.063]	1.949 [0.143]	$\beta > 0.05$	$\beta >> 0.01$	Checked
3. Rural Minor Arterial (4)	0.4971 [0.023]	5.384 [0.001]	$\beta > 0.01$	$\beta = 0.001$	Checked

B. Burr-Foster Q-Test (Unequal sample size)					Homogenity of Variance
Highway Category (No. of station or group)	Calculated q	Critical q		Remarks on β-level	
		$\beta = 0.01$	$\beta = 0.001$		
4. Rural Major Collector (3)	0.4938	0.4827	0.5543	$\beta = .01 - .001$	Checked

homogeneity of variance for Rural Minor Arterial was accepted at a β-level of 0.001. The Cochran "C values" were checked against critical "C values" (Appendix Table C.8 [48]) in the first two categories of highway with a β-level of 0.05 and for rural minor arterial with a β-level of 0.01.

The Burr-Foster critical q-value [3] shows β-levels for rural major collectors between 0.01 and 0.001. So, the homogeneity of variance was accepted for rural major collector at a β-level of 0.001, using the same reasons discussed above for rural minor arterial.

5.2.2 Normality

The normality of the four data tables for the four highway categories of Appendix A was analyzed by means of the Shapiro-Wilk test [39] for each station separately and after combining stations within a highway category. Because of the few stations in each category of highway, normality is not expected when stations in a highway category are analyzed together. But for each station separately, normality is an expected result. The result of this test is shown in Table 5.2. In this table, the small values of W with smaller β-level are significant, i.e., lead to rejection of the normality.

Table 5.2

Results of the Test for Normality

Highway Category	Station(s)	No. of Cases	Shapiro-Wilk W	B-level	Normality
1. Rural Inter-state	(i) All stations	26	0.9302	0.05 - 0.10	Checked
	(ii) 172A	13	0.8909	0.10 - 0.50	Checked
	(iii) 3070A	13	0.9227	0.10 - 0.50	Checked
2. Rural Principal Arterial	(i) All stations	39	0.8922	<0.01	Unchecked
	(ii) 68A	13	0.9161	0.10 - 0.50	Checked
	(iii) 173A	13	0.9254	0.10 - 0.50	Checked
	(iv) 254E	13	0.9227	0.10 - 0.50	Checked
3. Rural Minor Arterial	(i) All stations	52	0.8880	<0.01	Unchecked
	(ii) 25A	13	0.9003	0.10 - 0.50	Checked
	(iii) 301A	13	0.9753	>0.50	Checked
	(iv) 313A	13	0.8967	0.10 - 0.50	Checked
	(v) 262A	13	0.9171	0.10 - 0.50	Checked
4. Rural Major Collector	(i) All stations	37	0.8051	<0.01	Unchecked
	(ii) 47A	11	0.9167	0.10 - 0.50	Checked
	(iii) 59A	13	0.9143	0.10 - 0.50	Checked
	(iv) 5420A	13	0.8899	0.10 - 0.50	Checked

The test results for normality show that the Y's are normal at a β-level greater than 0.10 within each station location. Some of the stations satisfied the normality criterion with a β-level greater than 0.50. But the Y's of all the stations together under a highway category did not satisfy the normality criterion, except Rural Interstates at a β-level greater than 0.05. Different transformations (for example: square-root, log, $(Y_{max}^{0.5} - [Y_{max} - Y_i]^{0.5})$, etc.) were done on Y's to satisfy normality for each category of highway. But these transformations failed to satisfy normality, when the normality test was done on the transformed Y's.

The reason for nonnormality within a category of highway is the wide variation of Y's among the stations. The addition of stations would help to achieve normality. In case of Rural Interstates, the normality hypothesis was accepted at a β-level of 0.05.

5.2.3 Scattergram

Scatterplots of the dependent variable (AADT) against the independent variables are presented in Appendix B for the four highway categories and in Appendix D for the two stations -- 68A and 7047A. The scatterplots were prepared with the help of the

Statistical Package for the Social Sciences (SPSS) [39] to identify any apparent trends among the variables. The plots in Appendix B show the gaps and clusters among the stations. The addition of new count stations would help to remove or reduce such gaps and establish better statistical relationships. The plots in Figures D1 to D17 do not show any clusters, but these plots show a general linear trend. Slight departures from the linear trend are noticed in the Appendix D plots at years beginning with 1980. The plots of AADT vs. Gasoline Price are more scattered and thus indicate less correlation between these variables. Although clusters are found in several plots in Appendix B, a good linear trend is present (for example, Figures B1.8, B3.5, B4.4 etc.).

5.2.4 Conclusions from Preliminary Analysis

Homogeneity of variance tests, considering each station as a group, shows equal variances among the groups for each category of highway. The normality hypothesis is accepted for each station separately and for Rural Interstates as a group. The reason for normality of Y's for Rural Interstates is insignificant variation in Y's at its two stations. The main reason for nonnormality in the other

categories of highways is the wide variation or gap in Y's among the stations, i.e., an insufficient number of count stations for each category of highway. At the same time, due to fewer observations in each category of highway, sampling of data was not done. The normality assumption is an expected result for sampling cases when such kind of pooling is done. It is apparent that, with the installation of new stations that will eliminate the gaps in Y's, the Y's will tend to be normal. It is true that the Y's are not experimental and hence normality is possible only with counts of Y's between the gaps when useful transformations on Y fail to achieve normality. The normality test shows that analysis for each station separately will yield a better model than that for the combination of stations within a highway category.

It appears that normality tests with the available count stations do not support the idea of combining the stations within a category of highway. But it is also clear from these analysis that the normality of Y's for a category of highways is expected for a larger number of count stations and/or in sampling cases in pooled analysis. On the other hand, each station AADT data do confirm the normality assumption (See Table 5.2). The scatterplots for the

stations -- both separately and together -- do not show gross departures from a linear relationship in most of the cases for demographic and economic indicators. No recognizable pattern other than linear is noticeable in these plots. Scatterplots of gasoline price and time (Appendix Figures B1.2, B2.2, B3.2, B4.2, and B1.3, B2.3, B3.3, B4.3) are more scattered and indicate lower correlation with AADT.

In the next two sections (5.3 and 5.4), two types of analyses will be carried out. In Section 5.3, aggregate analysis combining all the stations within a category of highway is employed to develop an aggregate model for each category of highway. In Section 5.4, disaggregate analysis of each station separately is performed, and the resulting models will be location-specific.

5.3 Aggregate Analysis

In this section, models are developed for each of the six categories of highway. In the selection of variables, theoretical judgments, together with the results of statistical analyses, are taken into consideration. After developing the models, their performance is tested against the data for the stations that were not used in the development of model.

In aggregate analysis, the stations were pooled under a category of highway. But the data for stations clearly out of the range of values for most of the stations in its category were not used in the development of a model. From a statistical standpoint, it is wise to restrict prediction to the region of the X-space from which original data were obtained. In case of this aggregate analysis, the X-space becomes wide enough with respect to disaggregate analysis X-space. Aggregation of data also helps to increase the number of observations or cases.

5.3.1 Multiple Linear Regression Analysis

In this section, the results of some analyses are presented. Each analysis is discussed briefly, together with some interpretations and criteria for selection.

5.3.1.1 Correlation Matrix

The statistical analysis begins with the study of the correlation matrix for the various factors considered. Table 5.3 shows the correlation matrix for the four categories of highway. The SPSS [39] regression program was used to obtain the correlation matrix. The correlation coefficient (r) in this table shows the intercorrelation between the variables considered.

An important fact regarding this correlation coefficient is that, when independent variables are highly correlated, the regression coefficient of any independent variable depends on which other independent variables are included in the model. In the case of highly correlated independent variables, a regression coefficient does not reflect any inherent effect of the particular independent variable on the dependent variable, but only a marginal or partial effect, given whatever other correlated independent variables are included in the model. [15,19,37]

In general, when two independent variables are correlated between each other, intercorrelation or multicollinearity among them is said to exist [37]. The three important problems that arise when using the highly correlated variables are:

Table 5.3

Correlation Matrix (*)
(Aggregate Analysis)

(A) Rural Interstate

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	.575												
X2	.402	.680											
X3	.728	.853	.827										
X4	.265	.912	.520	.607									
X5	.538	.984	.757	.890	.899								
X6	.647	.229	.507	.619	-.169	.251							
X7	.825	.866	.779	.975	.605	.876	.636						
X8	.811	.860	.797	.974	.602	.874	.640	.993					
X9	.732	.855	.830	.999	.607	.890	.629	.978	.980				
X10	.725	.739	.626	.763	.497	.704	.625	.857	.847	.781			
X11	.575	.800	.865	.974	.580	.863	.594	.905	.911	.972	.676		
X12	.785	.859	.772	.972	.597	.870	.657	.988	.987	.979	.872	.923	
X13	.817	.861	.754	.974	.600	.873	.644	.993	.990	.979	.857	.914	.994

(B) Rural Principal Arterial

	Y	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11	X12
X1	.786												
X2	.275	.519											
X3	.398	.639	.827										
X4	.804	.878	.224	.259									
X5	.881	.938	.364	.429	.974								
X6	.633	.901	.543	.692	.667	.764							
X7	.405	.647	.779	.975	.247	.411	.717						
X8	.402	.642	.797	.974	.251	.415	.721	.993					
X9	.401	.640	.830	.999	.260	.429	.702	.978	.980				
X10	.354	.547	.626	.763	.191	.316	.697	.857	.847	.781			
X11	.373	.602	.865	.974	.260	.427	.654	.905	.912	.972	.676		
X12	.413	.641	.773	.972	.252	.416	.735	.987	.987	.979	.872	.923	
X13	.412	.643	.754	.974	.251	.414	.727	.993	.990	.979	.857	.915	.994

Table 5.3 (continued)

(C) Rural Minor Arterial

	Y	X1	X2	X3	X4	X5
X1	.800					
X2	.018	.307				
X3	.068	.383	.827			
X4	.907	.954	.123	.149		
X5	.853	.989	.241	.290	.986	
X6	.056	.416	.514	.646	.240	.345

(D) Rural Major Collector

	Y	X1	X2	X3	X4	X5
X1	.766					
X2	.178	.593				
X3	.164	.687	.818			
X4	.915	.901	.354	.341		
X5	.731	.954	.587	.618	.921	
X6	-.453	.121	.452	.568	-.163	.205

(*) For definition of variables, see Table 4.1;
Xi represents X_i, where i = 1 to 13.

1. Adding or deleting an independent variable changes the regression coefficients.

2. The extra sum of squares of regression associated with an independent variable varies depending upon which independent variables are already in the model.

3. The estimated regression coefficients individually may not be statistically significant, even though a definite statistical relationship exists between the dependent variable and a set of independent variables. These problems can also arise without substantial multicollinearity being present, but only under unusual circumstances, not likely to be found in practice.

The existence of multicollinearity does not invalidate a regression analysis, but neither is the absence of multicollinearity a validation of a particular regression model. Multicollinearity is also not a specification error [19]. The results of the correlation coefficients will play a role in the selection of variables for the model under development.

5.3.1.2 Stepwise Regression

The stepwise regression procedure is the most widely used automatic search method. It selects one variable at a time for entry into the model, until a desired subset of

variables is selected. The stepwise regression was carried out with the help of the SPSS package [39]. The summary of that analysis is shown in Table 5.4. The order in which variables entered into the regression model does not reflect their importance in the model [37].

In designing the regression statement, the four associated parameters (number of steps, F-value to enter [FIN], tolerance, and F-value to out [FOUT]) play important roles in the selection of variables for the models. Three cases were considered in using these parameters. In case A, all the parameters are default parameters. This case will allow most variables to enter the regression equation and will seldom force out a variable during the stepwise procedure. The selection of FIN, FOUT and tolerance level values in cases B and C allows more control by the analyst over variable selection. For cases B and C, FIN and FOUT have been computed using an F-table [37] of values $F(1-\alpha,1,n-p)$, where α is the associated level of significance, p is the expected number of terms in the regression equation (a value of 3 was used), and n is the number of cases or observations. FOUT was kept less than FIN. The calculated values of FIN and FOUT were shown in column 2 of Table 5.4. For the parameter "number of steps", a default parameter of twice the number of independent variables was used in all three cases. Since

Table 5.4

Stepwise Regression Summary (Aggregate Analysis)

Highway Category	Case (*) & Parameter	Step	Variable subscript: Entered	Variable subscript: Removed	F value	Significance level	Last-step b-coeff.	R Squared	Overall F (**)
RURAL INTERSTATE	Case A: Default Parameters	1	7		51.255	0.0	.0063	.681	51.255
		2	11		24.314	0.0	-12.8498	.845	62.681
		3	5		17.446	0.0	-4.3296	.914	77.470
		4	3		13.263	.002	441.6104	.947	93.866
		5	13		9.256	.006	8.9645	.964	106.406
		6	10		5.294	.030	-.0035	.972	108.593
		7	4		4.197	.055	.6672	.977	109.339
		8	12		2.545	.029	-16.6688	.980	104.201
		9	6		1.122	.305	.6764	.981	93.412
		10	1		.460	.506	.0090	.982	81.276
		11	8		.046	.834	.7645	.982	69.190
		12	2		.369	.554	-33.7152	.982	60.598
		Constant term					-820756.16		
	Case B: Alpha=.10 FIN=3.00 FOUT=2.90	1	7		51.255	0.0	.0457	.825	51.255
		2	11		24.314	0.0	-34.8261	.919	62.031
		3	5		17.446	0.0	-.0801	.956	77.471
		4	10		12.509	.002	-.0074	.973	91.601
		5	13		13.509	.001	7.7625	.984	119.696
		Constant term							
	Case C: Alpha=.05 FIN=4.35 FOUT=4.25	(SAME AS CASE B)							

Table 5.4 (continued)

Highway Category	Case (*) & Parameter	Step	Variable subscript Entered	Variable subscript Removed	F value	Significance level	Last step b-coeff.	R Squared	Overall F (**)
RURAL PRINCIPAL ARTERIAL	Case A: Default Parameters	1	5		128.457	0.0	4.0248	.776	128.455
		2	4		12.498	.001	-1.1233	.834	98.418
		3	11		34.666	0.0	21.9403	.916	128.206
		4	6		11.869	.002	-.3541	.964	126.587
		5	10		24.297	0.0	.0828	.964	175.521
		6	2		15.684	0.0	-89.3566	.976	213.965
		7	1		18.672	0.0	-.2149	.985	287.347
		8	12		7.785	.009	2.2668	.988	387.432
		9	3		1.736	.198	-943.8548	.989	288.171
		10	7		1.432	.241	.0051	.989	256.055
		11	13		.512	.292	-2.6588	.989	228.764
		12	8		1.159		.0087	.990	211.834
		Constant term					1010953.7		
	Case B: Alpha=.10 FIN=2.90 FOUT=2.80	1	5		128.457	0.0	3.9918	.776	128.455
		2	4		12.498	.001	-1.0506	.834	98.418
		3	11		34.666	0.0	-17.2873	.916	126.206
		4	6		11.069	.002	-.2659	.964	126.587
		5	10		24.297	0.0	.0834	.964	175.521
		6	2		15.684	0.0	-37.5672	.976	213.965
		7	1		18.672	0.0	-.3098	.985	287.347
		8	12		7.785	.049	5.4797	.988	387.432
		Constant term					711.7891		
	Case C: Alpha=.05 FIN=4.20 FOUT=4.10	(SAME AS CASE B)							

Table 5.4 (continued)

Highway Category	Case (*) & Parameter	Step	Variable subscript Entered	Variable subscript Removed	F value	Significance level	Last step b-coeff.	R Squared	Overall F (**)
RURAL	Case A: Default Parameters	1	4		233.176	0.0	.6129	.823	233.175
		2	5		26.047	0.0	-1.9851	.865	168.016
		3	3		103.742	0.0	200.2018	.964	422.743
		4	1		8.609	.005	.1113	.969	369.407
		5	6		5.571	.023	-.0564	.973	325.382
		6	2		2.946	.093	-12.7288	.974	283.114
		Constant term					-4081161.6		
MINOR	Case B: Alpha=.10 FIN=2.05 FOUT=2.75	1	4		233.175	0.0	.5453	.823	233.175
		2	5		26.047	0.0	-1.4905	.884	168.016
		3	3		103.742	0.0	213.5501	.963	422.743
		4	2		4.405	.041	-17.7518	.967	349.650
		5	6		5.609	.030	-.0541	.978	296.768
		Constant term					-418772.2		
ARTERIAL	Case C: Alpha=.05 FIN=4.10 FOUT=4.00	(SAME AS CASE B)							
RURAL	Case A: Default Parameters	1	4		188.862	0.0	.5983	.837	188.862
		2	6		47.491	0.0	-.0492	.932	233.367
		3	1		2.818	.103	-.0569	.937	164.834
		4	5		1.820	.187	-1.0948	.941	127.151
		5	3		3.125	.087	152.0768	.946	109.102
		6	2		.302	.587	-6.0766	.947	88.921
		Constant term					-380813.9		
MAJOR	Case B: Alpha=.10 FIN=2.95 FOUT=2.80	1	4		188.862	0.0	.2564	.837	188.862
		2	6		47.491	0.0	-.1979	.932	233.367
		Constant term					-4953.47		
COLLECTOR	Case C: Alpha=.05 FIN=4.20 FOUT=4.10	(SAME AS CASE B)							

(*) A. Default Parameters:

(i) Max. No. of Steps = 2 * No. of Independent Variables [For All Cases]

(ii) FIN = .01, FOUT = .005 [For Case A]

(iii) Tolerance level = .001 [For Case A]

B. Tolerance level = .01 [For Case B & Case C]

(**) Overall Significance = 0.0

the degrees of freedom associated with Mean Squared Error (MSE) vary, depending on the number of X variables in the model, and since repeated tests on the same data are undertaken, fixed F-limits for adding or deleting a variable have no precise probabilistic meaning [37]. MSE is defined as Sum Squared Error (SSE) -- sum of squared of deviations around the regression line or plane -- divided by its degrees of freedom, n - p. A minimum tolerance of 0.01 was used in case B and case C to guard against the entry of a variable that is highly correlated with other X variables already in the model. The tolerance is defined as $1 - R_K^2$, where R_K^2 is the coefficient of multiple determination when X_K is regressed on the other X variables in the regression model. The tolerance specification of 0.01 provides that no variable is to be added to the model if it has a coefficient of multiple determination with the other X variables already in the model that exceeds 1 - .01 = 0.99 or that would cause the R_K^2 for any variable in the model to exceed 0.99.

5.3.1.3 C_P-statistic in All Possible Regression

The C_P-statistic, R^2, etc. for a reasonable number of subsets of variables were calculated with the help of an program "DRRSQU" [42]. Some of those C_P and R^2 values are shown in Table 5.5.

<u>Table 5.5</u>

Selected C_p & R-Squared in All Possible Regression
(Aggregate Analysis)

Highway Category	Subscripts of Variables in Equation	C_p Values in same order	R-Squared Values in same order	P
RURAL INTERSTATE	7 13 8 12 9 10	213.7,223.2 238.4, 264.4,321.1, 329.1	.681, .665, .658, .613,.536, .525	2
	3 11 9 11, 7 11, 2 7, 2 8, 7 9	72.9, 73.6, 94.6, 106.4,119.2, 120.7	.874, .873, .845, .829,.824, .81	3
	1 9 11, 2 5 7, 1 2 7, 2 5 8, 2 4 7	27.5, 42.6, 52.7, 56.7, 59.2	.908, .918, .904, .893, .896	4
	....	......	......	...
	Subscripts of all variables	14	.984	14
RURAL PRINCIPAL ARTERIAL	5 4 1 6 12	559.6,903.7,981.4, 1557.82178.3	.776, .647, .613, .481, .171	2
	4 5, 1 5, 5 7 5 6, 5 9, 4 7, 4 6, 4 9	483.4 524.3,555.8, 557.1,559.9,785.4, 791.9, 891.3	.834, .798, .779 .776, .777, .692, .690, .686	3
	4 5 11, 3 4 5, 1 4 5, 4 5 9 2 4 5, 4 5 6, 4 5 8, 4 5 7	196.8,227.1,227.3, 234.9,239.7,288.8, 298.8, 395.4	.917, .903 .903, .906, .898, .883, .876, .873	4
	....	......	......	...
	Subscripts of all variables	14	.995	14

Table 5.5 (continued)

Highway Cate-gory	Subscripts of Variables in Equation	C_p Values in same order	R-Squared Values in same order	P
RURAL MINOR ARTE-RIAL	4, 5, 1, 3, 6	259.8, 420.1, 578.7, 1667.7, 1696.7	.820, .727, .641, .005, .003	2
	4 5, 1 4, 4 6, 2 4, 1 5, 2 5, 1 6, 1 2	155.8, 178.4, 213.3, 246.4, 287.8, 364.9, 419.1, 481.2	.885, .871, .851, .832, .809, .764, .733, .698	3
	2 4 5, 1 4 5, 4 5 6, 1 4 6, 2 4 6	132.4, 150.5, 157.0, 178.4, 215.1	.899, .888, .885, .872, .851	4
	....	......	......	...
	Subscripts of all variables	7	.974	7
RURAL MAJOR COL-LECTOR	4, 1, 5, 6, 3	58.7, 199.9, 229.8, 414.9, 515.5	.837, .587, .534, .205, .027	2
	4 6, 4 5, 5 6, 1 6, 2 4, 2 5	7.26, 13.46, 18.18, 31.46, 46.87, 178.3	.932, .921, .913, .889, .862, .629	3
	3 4 5, 1 4 6, 3 4 6, 1 4 5, 2 4 6, 4 5 6	2.40, 6.25, 6.76, 8.43, 9.12, 9.19	.944, .937, .937, .934, .932, .932	4
	....	......	......	...
	Subscripts of all variables	7	.947	7

The C_p-criterion is concerned with the total mean squared error (MSE) of the n fitted values for each of the various subset regression models. When the C_p values for all possible regression models are plotted against P, those models with little bias will tend to fall near the line $C_p = P$ [15]. Models with substantial bias will tend to fall considerably above this line. In using the C_p-criterion, the subsets of X variables for which (1) C_p value is small and (2) the C_p value is near P, are considered for the model. Sets of X variables with small C_p values have a small total mean squared error, and when the C_p value is also near P, the bias of the regression model is small. It may sometimes occur that the regression model based on the subset of X variables with the smallest C_p value involves substantial bias. In that case, one may at times prefer a regression model on a somewhat larger subset of X variables for which the C_p value is slightly larger, but which does not involve a substantial bias component. Thus, one should look for a regression with a low C_p value about equal to P. When the choice is not clear-cut, then it is a matter of personal judgment whether one prefers a biased equation or an equation with more parameters. Draper and Smith [15] recommend the use of the C_p-statistic in conjunction with the stepwise method to choose the best equation. Some statisticians suggest that all possible regression models with a similar number of X variables to the number in the

stepwise regression solution be fitted subsequently to investigate which subset of X variables might be best [37].

The final selection of the model variables will be aided by residual analyses. Information gained by these analyses, together with the investigator's knowledge about the phenomenon under study, will be helpful in choosing the final regression model to be employed [37].

5.3.2 Preliminary Screening of Candidate Variables

The screening of variables was not confined to statistical analysis. Judgment regarding the questions listed in Table 5.6 was considered while preparing data tables prior to regression analysis. No screening of variables was done at that stage, however. The initial inclusion of a large number of variables in the models for Rural Interstates and Rural Principal Arterials is justified by the fact that the omission of essential variables may produced biased estimates while the inclusion of large number of variables does not [19]. The basic questions in Table 5.6 will again be reviewed in the selection of the variables. The goals of the analysis that should be met in this selection process are shown in Table 5.7. How these goals are considered for each category of highway is demonstrated in the following sections.

Table 5.6

Some Fundamental Criteria for Variable Selection [15,19]

1. Are the proposed variables fundamental to the problem?
2. Availability of data (variables).
   - (a) Are annual data available?
   - (b) Are historical data available?
   - (c) What is the most recent year of data?
   - (d) Will data be available in future?
3. Cost to obtain the data.
4. How reliable is the data?

Table 5.7

Goals of the Analysis

1. The final equations should explain more than 50% of the variation ($R^2 > 0.50$).
2. The C_p value will be lowest and near to P.
3. The number of predictor variables should be adequate for each model (*).
4. The selection will respond well to the questions of Table 5.6.
5. All estimated coefficients in the final model should be statistically significant at an alpha-level of 0.05 or 0.10.
6. There should be no discernible patterns in the residuals.

(*) As a general rule, there should be about ten complete sets of observations for each potential variable to be included in the model; e.g., if it is believed that the final practical predictive model should have four X-variables plus a constant, then there should be at least forty sets of observations (n = 40) [15].

5.3.2.1 Rural Interstates

The correlation matrix in Table 5.3(A) shows that both X_1 and X_4 have moderate correlation with Y (r_{Y,X_1} = 0.575 and r_{Y,X_4} = 0.265), but the correlation coefficient between these variables are quite high ($r_{1,4}$ = 0.912). The variables X_7 to X_{13} are highly intercorrelated with each other. Any one of them -- as opposed to all of them -- is eligible to explain Y and to lessen multicollinearity.

The case A stepwise regression with default parameters includes almost all the variables, but the sign of b-coefficients in the cases of X_5, X_{10}, X_{11} and X_{12} is negative, which is contrary to the expected positive sign indicated in scatterplots (Figures B1.5, B1.10, B1.11 and B1.12) for the respective variables. The reason for this unexpected result is the high intercorrelation between some of the variables. The case B and case C stepwise regressions entered X_5, X_7, X_{10}, X_{11} and X_{13} into the equation with negative b-coefficients X_5, X_{10} and X_{11} with R^2 of 0.984. The best subset according to the C_P-criterion has too many variables. Furthermore, the correlation coefficients among the variables are in some cases higher than 0.90.

Considering all the points discussed above, the good subsets at P = 2, 3 and 4 in Table 5.5 and X_2, X_9 at P = 3

with R^2 = 0.673 and X_5, X_7 and X_{11} at P = 4 with R^2 = 0.914 will be further analyzed to make the final selection from them.

5.3.2.2 Rural Principal Arterials

The correlation matrix Table 5.3(B) shows that X_1, X_4, X_5 and X_6 are highly correlated with Y ($r > 0.633$). The gasoline price (X_2) has the lowest correlation with Y ($r = 0.275$). X_1, X_4 and X_5 are highly correlated among themselves ($r > 0.878$), which argues for the use of only one of these variables to avoid multicollinearity in the resulting model. The variables X_7 to X_{13} are also highly intercorrelated and only one of these should be selected to avoid multicollinearity.

The case A stepwise regression with default parameters entered almost all variables with negative signs in b-coefficients in X_1, X_3, X_4, X_6 and X_{13}. (See Table 5.4). These negative signs are contrary to the expected positive signs indicated by the scatterplots (Figures B2.1, B2.3, B2.4, B2.6, and B2.13). The reason for these negatively signed b-coefficients is a high degree of intercorrelation among some independent variables, as shown in Table 5.3(B). So, the case A stepwise regression choice will not be further analyzed if other choices in Table 5.5 avoid this problem.

The case B and case C stepwise regressions entered eight variables out of thirteen with negative b-coefficients X_1, X_4, X_6 and X_{11} and with R^2 of 0.988. The R^2 of 0.990, in the case A stepwise regression with twelve variables, increased only a negligible amount with respect to the eight variables in the equation for the case B and case C stepwise regressions.

In choosing the C_P and R^2 values in Table 5.5, judgment regarding questions of Table 5.6 and correlation coefficients values between the variables were taken into consideration because there was a large number of subsets that could be considered. For example, X_1, X_4, and X_5 are highly intercorrelated ($r > 0.878$), and anyone from these is considered, because X_1, X_4, X_5 and X_6 are almost equally correlated with Y ($r \simeq .800$). The best subset according to the C_P-criterion has too many variables. Considering all the points discussed above together with the goals of analysis of Table 5.17, the good subsets of variable sets at P = 2 and P = 3 in Table 5.5 will be further analyzed to make the final selection from them.

5.3.2.3 Rural Minor Arterials

The correlation matrix Table 5.3(C) shows that X_1, X_4 and X_5 have almost equal correlation with Y ($0.800 \leq r \leq 0.907$). The variables X_1, X_4, and X_5 are highly intercorrelated ($r > 0.954$).

The case A stepwise regression with default parameters includes all variables with an R^2 of 0.974 (see Table 5.4). However, an R^2 of 0.823 was obtained with only X_4 at step 1. The case B and case C stepwise regressions enter all the variables except X_1 with an R^2 of 0.970. The b-coefficients of X_2,X_5 and X_6 are negative. The negative coefficient of X_2 (gasoline price) is an expected result. The reason for the negative coefficients of X_5, and X_6 is its high correlation with other variables in the model (for example, $r_{1,5}$ = 0.989, $r_{3,6}$ = 0.646). With X_5 and X_6 alone in the equation, the sign of its b-coefficient was positive. The best subset according to the C_p-criterion has too many variables. Moreover, the subset with more than one variable usually has high correlation between the variables (for example, $r_{4,5}$= 0.986).

Considering all the points discussed above and the criteria of Table 5.6, the following subsets of variables were kept for the final selection process:

1. X_5
2. X_4
3. X_1
4. X_4, X_6
5. X_5, X_6
6. X_2, X_4
7. X_2, X_5

8. X_4, X_5
9. X_1, X_4
10. X_2, X_4, X_5
11. X_1, X_4, X_5
12. X_4, X_5, X_6

All these choices will provide an R^2 of at least 0.641.

5.3.2.4 Rural Major Collectors

The correlation matrix Table 5.3(D) shows that X_1, X_4, and X_5 have good correlation coefficients with Y of 0.766, 0.915 and 0.866, respectively. County employment (X_6) and AADT (Y) are negatively correlated, which is not the expected relationship, so the selection of variable X_6 will not be considered unless supported by other analyses. The variable X_3 has the lowest correlation coefficient with Y ($r = 0.164$). X_1, X_4 and X_5 are highly correlated among each other ($r \geq 0.731$).

Table 5.4 shows that the case A stepwise regression with default parameters includes all the variables with an R^2 of 0.947. However, an R^2 of 0.837 was obtained with only X_4 at step 1. The case B and case C stepwise regressions select the variables X_4 and X_6 with an R^2 of 0.932. The inclusion of other variables in the case A stepwise regression increased R^2 by only a small amount. The b-coefficients of X_1 and X_5 in case A and X_6 in all cases are negative. The negative coefficient of X_2

(gasoline price) is an expected result. So, the case A stepwise regression choice will not be further analyzed, since other choices avoid the problems associated with it.

The C_p values in Table 5.5 show that the variable set X_3, X_4 and X_6 at P = 4 is the best selection, with C_p of 2.40 and R^2 of 0.944. But the selection of X_4 and X_6 at P = 3, with C_p = 7.26 and R^2 = 0.932, is the result of stepwise regression in cases B and C. The variable sets {X_1, X_4 and X_6} and {X_3, X_4 and X_6} at P = 4, with C_p of 6.25 and 6.76, respectively, are good for further analysis. Note that X_4 has high correlation with X_5 (r = 0.921). And X_4 has negative correlation with X_6 (r = -0.163), which is not an expected result.

Considering the questions of Table 5.6 and the results of the C_p-criterion, correlation matrix and stepwise regression, the following subsets were kept for final selection process:

1. X_4
2. X_1
3. X_5
4. X_4, X_6
5. X_5, X_6

6. X_4, X_2

7. X_5, X_2

These choices have R^2 of at least 0.534.

5.3.2.5 Summary of Preliminary Screening Process

The R^2 value, C_p-criterion, stepwise regression, correlation coefficients among variables, and the questions in Table 5.6 were taken into consideration in the screening of variables in the preliminary selection phase. The combination of these criteria, discussed separately under each category of highway, resulted in some good subsets of variables from which to make the final selection. The preliminary screening reduces much work in further analysis by considering only the good choices that result from it. In this screening process, the first four goals of Table 5.7 were taken into consideration. Subjective judgment also was made because it was not always possible to meet all four of those goals at the same time.

5.3.3 Final Selection of Variables

In the final selection, the goals of the analysis in Table 5.7 were considered together to find the best subset of variable(s) from the preliminary choices for each highway category. Goals 1 to 4 in Table 5.7 were taken into consideration in preliminary choices. Final selection of candidate variables from preliminary choices was done later through the careful examination of all criteria except the residual analysis and hypothesis testing concerning b-coefficients. The ith residual, denoted by e_i, is the difference between the observed value Y_i and the corresponding fitted value $\hat{Y}_i$ (i.e., $e_i = Y_i - \hat{Y}_i$). Residual analysis and testing concerning regression coefficients were carried out in the final selection. The final selection was then used to build the model. The variables' coefficients were scrutinized using the following three questions [15]:

1. Are the coefficients reasonable?

   The least squares regression coefficients are adjusted for other variables in the regression. Thus, the regression coefficients may attempt to predict the response by changing only one variable, using its coefficient to decide how much to change it. If all the estimated coefficients are independently estimated, this may do little harm.

However, when the predictor variables are highly correlated and the estimated coefficients are also correlated, reliance on individual coefficients can be dangerous. A check can also be made to see if individual coefficients are directionally correct. For example, if X_1 is number of vehicle registrations and Y is the AADT, then b_1 (the b-coefficient corresponding to X_1) should be positive. This question was examined by checking the positive or negative sign of coefficient with that of the expected sign.

2. Is the equation plausible?

Are the appropriate variables in the equation, and are any obvious variables missing? This question was considered in the residual analysis on final selection to see if any important variable was missed and by examining the first, third and fourth questions in Table 5.6.

3. Is the equation usable?

The final model will contain a set of variables that can be used for predicting response variable(s) (in this case, AADT). This question was considered through the variable selection process by considering the second question in Table 5.6 regarding the future value of the variable.

In the final selection of variable(s) for the model's equation, the criteria of establishing high R^2 was not considered exclusively. Because R^2 is a relative quantity, it indicates how large the Regression Sum of Squares (SSR -- sum of the squares of the deviations of the fitted regression values around mean) is relative to the Total Sum of Squares (SST -- sum of squares of total deviations around mean), where SST is fixed and does not depend on Y. SST = SSR + SSE, where SSE is the Error Sum of Squares or residual sum of squares -- sum of squares of the deviations around regression line or plane. In some situations, data may be quite variable and a large R^2 may not indicate a very good fit. In more controlled situations, a relatively small R^2 may indicate a rather good fit [19]. The value of R^2 can only increase if the number of predictor variables increases. Consequently, R^2 is always the maximum for the full set with all predictor variables. So maximizing R^2 cannot really be the sole selection criterion. However, one can subjectively choose a subset of predictor variables that gives a good value of R^2, such that using any additional predictor variables results in only a marginal improvement in R^2. The residual patterns were always examined on the final selection to accept the final selection for building the model.

5.3.3.1 Regression on Preliminary Choices and Final Selection

Regression on the preliminary choices was done with the help of the SPSS package [39]. The summary of that analysis is shown in Table 5.8 for all four categories of highway. The magnitude of b-coefficients and their inconsistency with reference to sign is shown in Table 5.8.

(1) Rural Interstates

Table 5.8 shows inconsistency in the b-coefficients in some of the preliminary choices. The more variables in the model, the more costly and complex it becomes to implement and maintain. If the model is restricted to variables without inconsistency in their coefficients and judgment is applied to the questions in Table 5.6, then the following choices are eligible for the final selection:

1. X_7
2. X_8
3. X_9
4. X_2, X_7
5. X_2, X_8
6. X_7, X_9

Inconsistency in regression coefficients is due to multicollinearity. It was mentioned in Section 5.3.1.1 that this multicollinearity does not invalidate the

Table 5.8

Multiple Linear Regression Summary on Preliminary Choices (Aggregate Analysis)

Highway Category	Variable Subscripts in Eqn.(*)	b-coefficient in same order	Inconsistencies in b's(**)	R-Squared	Overall F(***)
	7	.0034	---	.681	51.255
	8	.0154	---	.656	45.275
	9	.0111	---	.536	27.705
RURAL					
	2,7	-148.405, .0054	---	.829	55.714
	2,8	-163.195, .0255	---	.824	53.641
	2,9	-160.829, .0135	---	.673	2[illegible].645
	7,9	0.0183, -.0261	-b9	.819	48.986
	7,11	0.0070 -29.696	-b11	.845	62.681
INTER-					
STATE	1,2,7	-.1869 -146.626, .0073	-b1	.904	69.373
	2,5,8	-140.087, -.0640, .0336	-b5	.893	65.206
	2,4,7	-135.718, -.1593, .0061	-b4	.896	62.097
	5,7,11	-.0979, 0.0062, -22.0067	-b5,-b11	.914	73.479
RURAL	5	.0465	---	.776	128.457
	4	.3634	---	.647	67.308
	1	.3105	---	.618	59.786
PRINCI-	4,7	.3366 .00039	---	.692	40.484
PAL	4,8	.3391, .00443	---	.695	45.017
	4,9	.3393, .00341	---	.666	39.364
	5,(7)	.0262, .00023	---	.779	63.289
	5,(8)	.0266, .00031	---	.778	63.100
ARTERIAL	5,(9)	.0353 .00045	---	.777	62.716

Table 5.8 (continued)

Highway Category	Variable Subscripts in Eqn.(*)	b-coefficient in same order	Inconsis- tencies in b's(**)	R-Squared	Overall F(***)
	5	.3157	---	.727	133.136
RURAL	4	.1032	---	.823	233.175
	1	.1157	---	.641	89.123
MINOR	4,6	.1076, -.1162	-b6	.851	146.278
	5,6	.3583, -.1829	-b6	.791	93.028
ARTERIAL	(2),4	-15.982, .1045	---	.832	121.616
	2,5	-33.502, .3335	---	.764	79.470
	4,5	.2675, -.543	-b5	.865	183.815
	1,4	-.1056, .1825	-b1	.871	165.831
	2,4,5	28.8886, .3462, -.8169	+b2,-b5	.899	142.197
	(1),4,5	.1088, .3244, -.9813	-b5	.888	127.451
	4,5,6	.2653, -.5351, -.3819	-b5,-b6	.865	122.806
RURAL	4	.2715	---	.837	188.862
	1	.1991	---	.587	49.693
MAJOR	5	.7122	---	.534	48.056
	4,6	.2564, -.1979	-b6	.932	233.367
COLLEC-	5,6	.8379, -.3938	-b6	.913	177.805
TOR	4,2	.2892, -34.5641	---	.862	106.830
	5,2	.9292, -78.3612	---	.629	28.772

(*) 95% confidence interval of the b-coefficient(s) for the variable(s) enclosed in bracket inclues zero.

(**) '+' and '-' sign with b-coefficient(s) is inconsistent with the expected result.

(***) Overall significance=0.00

regression analysis. The variable X_3 (year) has been dropped because it is believed that its effect is reflected in other variables and because year as a variable will always increase, while AADT may decrease with year. X_{11} (consumer price index) has been dropped because it is a US city average data and its increasing pattern has no theoretical bearing on the observed upward trend of AADT.

The important test statistics evaluated earlier in this chapter for the six candidates are summarized in Table 5.9.

Table 5.9

Summary Statistics of Choices for Final Selection (Rural Interstate)

Choice Number	Variable Subscripts in Equation	C_p	P	R Squared	Inconsistencies in b's
1	7	213.7	2	.681	---
2	8	230.4	2	.658	---
3	9	321.1	2	.536	---
4	2, 7	106.4	3	.829	---
5	2, 8	110.2	3	.824	---
6	7, 9	120.7	3	.810	---

These choices contain no inconsistencies in the regression coefficient. In Table 5.9, the R^2 and C_p values for choices 1 and 2 are almost equal. The R^2 values for

choices with two variable are not higher than that for choices with one variable. So, choice 2 of Table 5.9 with X_8 only is taken as the final selection for further analysis for Rural Interstates.

(2) Rural Principal Arterials

All the choices in Table 5.8 do not exhibit any inconsistency in regression coefficients but 95 percent confidence interval of some regression coefficients includes zero. The choices with zero in the 95 percent confidence interval of regression coefficients will not be considered for final selection. Choices with one variable in their equations have R^2 values of 0.618 to 0.776. There are choices with 2 variables in an equation without the inconsistency in b-coefficients and with R^2 greater than the choices with one variable. The following choices emerged as candidates for the final selection:

1. X_5
2. X_4
3. X_1
4. X_4, X_7
5. X_4, X_8
6. X_4, X_9

Regarding the questions of Table 5.6, it is apparent that all the variables in these final candidates are eligible to build the model.

The important test statistics evaluated earlier in this chapter for the above three candidates are summarized in Table 5.10.

Table 5.10

Summary Statistics of Choices for Final Selection (Rural Principal Arterial)

Choice Number	Variable Subscripts in Equation	C_P	P	R Squared	Inconsistencies in b's
1	5	559.6	2	.776	---
2	4	903.7	2	.647	---
3	1	981.4	2	.618	---
4	4, 7	785.4	3	.692	---
5	4, 8	791.9	3	.69	---
6	4, 9	801.3	3	.686	---

The b-coefficients of X_4 and the R^2 values for the last three candidates are approximately the same (See Tables 5.8 and 5.10). All the choices of Table 5.10 will provide somewhat biased estimation with respect to the C_P-criterion. Considering the questions of Table 5.6, X_8 is preferable to X_7 or X_9, because annual historical data of state population (X_8) is available, but historical data of state household (X_9) is computed based on data on X_8. So, the data on X_8 are more reliable and less costly than that on X_9. Future data on X_7 are not available. Thus, the

fifth choice in Table 5.10 (variables X_4 and X_8) is the final selection for further analysis for Rural Principal Arterials.

(3) Rural Minor Arterials

Table 5.8 shows inconsistency in the b-coefficient in some of the preliminary choices. The 95 percent confidence interval of b-coefficients of some of the variables includes zero. At the same time, R^2 in last three choices in Table 5.8 does not increase much with respect to earlier choices.

The important statistics evaluated earlier in this chapter for the remaining four choices of Table 5.8 are summarized in Table 5.11.

Table 5.11

Summary Statistics of Choices for Final Selection (Rural Minor Arterial)

Choice Number	Variable Subscripts in Equation	C_p	P	R Squared	Inconsistencies in b's
1	4	259.8	2	.823	---
2	5	428.1	2	.727	---
3	1	1578.7	2	.641	---
4	2, 5	364.9	3	.764	---

The first two choices of Table 5.11 are better than the other choices. The first choice has the largest R^2 (0.823) among the four candidates in Table 5.11 but is very close to the second choice. The variables X_4 and X_5 have future values available. Thus any of the first two choices in Table 5.11 is equally good for making the final selection for Rural Minor Arterials. The variable X_5 is being selected arbitrarily as the final selection for further analysis.

(4) Rural Major Collectors

Table 5.8 presents inconsistency in the b-coefficient for X_6 for the preliminary choices $\{X_4, X_6\}$ and $\{X_5, X_6\}$, respectively. The choice with X_4 has the largest R^2 and lowest C_p among all the choices with one variable. The choices with two variables without inconsistency in regression coefficients do not provide significant increase in R^2 with respect to the one-variable choices (see Table 5.8). Thus, the variable X_4 is the final selection for further analysis for Rural Major Collectors.

5.3.3.2 Graphic Residual Analysis on Final Selections

The residual plots shown in Appendix C were generated by the BMDP package [47]. The plots were done to check the aptness of each model. The ith residual, denoted by e_i, is the difference between the observed value Y_i and the corresponding fitted value $\hat{Y}_i$ (i.e., $e_i = Y_i - \hat{Y}_i$).

Figures C1.1 to C1.4 are the plots of residuals against predicted AADT, Figures C2.1 to C2.4 are the plots of residuals against the final selected predictor variables, Figures C3.1 to C3.6 are the normal probability plots of residuals (the residuals against their expected values under normality) and Figures C4.1 to C4.4 are the plots of residuals against year for the four categories of highway. In Figures C1.1 through C2.4 and C4.1 through C4.3, the number of points plotted at each position is printed.

The normal probability plots (Figures C3.1 to C3.4) fall reasonably close to straight lines, suggesting that the error terms are approximately normally distributed. A slight departure is noticed in the case of the normal probability plot for Rural Minor Arterials (Figure C3.3). It is believed [37] that this small departure from normality will not create any serious problems.

The plots of residuals against the fitted response variable and predictor variables, Figures C1.1 to C2.4,

indicate no ground for suspecting the appropriateness of the linearity of the regression function or constancy of the error variance. The clustering of residuals in some cases is the effect of combining the stations in the analysis. It is believed that a greater number of stations will remove these clustering patterns. There are no suggestions in any of these plots that systematic deviation from the fitted response plane (in case of more than one variable in the equation) or line (in case of one variable in the equation) is present. The error variance varies in some of these plots with the level of Y and X's, but this variation does not exhibit any gross departure. This slight variation with Y and X's level is the result of pooling data from stations in a particular category of highway. These residual plots against Y and X's do not indicate the presence of any outlier. In a residual plot, outliers are the points that lie far beyond the scatter of the remaining residuals, perhaps 4 or more standard deviations from zero [37].

Residual plots were also generated against variables not included in the model, to check whether some key independent or predictor variables could provide important additional descriptive and predictive power to the model. One such variable is the Year (X_3), which has not been included in any model. The plots of residuals against X_3, shown in Figures C4.1 to C4.4, do not indicate any

correlation between the error terms over time, since the residuals are random around the zero line. Thus, it is confirmed that the appropriate variables are included in the model and no additional variable will provide significant power to the model.

5.3.3.3 Testing Hypothesis Concerning Regression Coefficients

The F-test for the regression relation explains whether the variables in the model have any statistical relation to the dependent variable. The hypothesis is

$$H_0 : \beta_1 = \beta_2 = \ldots\ldots\ldots = \beta_{P-1} = 0$$

$$H_a : \text{all } \beta_k \; (k = 1, \ldots\ldots, P-1) \neq 0;$$

The test statistic is given by $F^* = \frac{MSR}{MSE}$. A sum of squares divided by its associated degrees of freedom is called a Mean Square (abbreviated MS), Regression Mean Square (denoted by MSR) is $\frac{SSR}{P - 1}$ and Error Mean Square (denoted by MSE) is $\frac{SSE}{n - P}$. The terms SSR and SSE have been defined earlier in Section 5.3.3.

If $F^* \leq F(1-\alpha, P-1, n-P)$, then H_0 holds and indicates that the variables in the model do not have any statistical relation to the dependent variable. Larger values of F^* lead to conclusion H_a. Table 5.12 shows the result at α-levels of 0.05 and 0.10. The test results conclude the hypothesis H_a (i.e., the relationships among the variables in the models exist) and H_a cannot be

Table 5.12

Overall F-tests for Aggregate Analysis

Highway Category	Variable Subscripts for Full Model	F^*	df_R , df_E (*)	α	Is H_3 true for	
					$\alpha = 0.05$?	$\alpha = 0.10$?
1. Rural Interstate	8	46.275	1, 24	<.001	Yes	Yes
2. Rural Principal Arterial	4, 8	40.017	2, 36	<.001	Yes	Yes
3. Rural Minor Arterial	5	133.136	1, 50	<.001	Yes	Yes
4. Rural Major Collector	4	180.062	1, 35	<.001	Yes	Yes

(*) df_R = degrees of freedom for Regression.
df_E = degrees of freedom for Error.

rejected at an α-level of as low as 0.05. Hence, the regression relationships listed in Table 5.12 exist.

To test the significance of each variable ($H_0 : \beta_k = 0$; $H_a : \beta_k \neq 0$ for $1 < k < P-1$) and each subset with more than one variable ($H_0 : \beta_1 = \ldots = \beta_j' = 0$; H_a : all $\beta_j \neq 0$ for $1 < j < P-1$), a general linear test [37] was employed. The applicable F-statistic is shown in equation 5.2.

$$F^* = \frac{\dfrac{SSE\ (R) - SSE\ (F)}{df_R - df_F}}{\dfrac{SSE\ (F)}{df_F}} \qquad (5.2)$$

where,

F^* = F statistic,

SSE (R) = Error Sum of Squares for the Reduced model,

SSE (F) = Error Sum of Squares for the Full model,

df_R = degrees of freedom of the Reduced model, and

df_F = degrees of freedom of the Full model.

The reduced model was obtained by dropping the element(s) to be tested from the full model under H_0. Table 5.13 shows the summary of the results obtained at α-levels of 0.05 and 0.10. The test results show that when variables are dropped from the model, there still exist regression relationships. The hypothesis H_a cannot be rejected at an α-level as low as 0.05 and each variable in the model has

Table 5.13

Partial F-tests for Aggregate Analysis

Highway Category	Variable Subscripts for		df_R , df_F (*)	F^*	α	Is H_a true for	
	Full Model	Reduced Model				α =.05?	α =.10?
1. Rural Interstate (**)							
2. Rural Principal Arterial	4,8	4 8	37, 36 37, 36	4.963 61.223	.025-.05 ≤.001	Yes Yes	Yes Yes
3. Rural Minor Arterial (**)							
4. Rural Major Collector (**)							

(*) df_R = degrees of freedom for SSE for Reduced Model and

df_F = degrees of freedom for SSE for Full Model.

(**) It has only one variable in Full Model.

a significant influence at a level of significance 0.05.

5.3.4 Model Development and Performance

The final regression equations are presented in Table 5.14, along with the R^2 values, overall F values, t-statistics and elasticities. The elasticities shown in this table were obtained from the output of Multiple Linear Regression on final selected variables computed according to equation 3.4 (Chapter 3). Not all the conditions of Table 5.7 have been met in all equations of Table 5.14. However, the equations that resulted from the specified criteria of Table 5.7 are the best possible, considering all the limitations. The equations in Table 5.14 use variables that are believed to be easily available from a variety of sources for both historical and future trends. Each of the variables is significant at the 95 percent confidence level. The equation for Rural Interstates has the lowest R^2 (0.658) and thus explains only 65.8 percent of the total variability of AADT by the use of variable, X_8. The equations for rural principal arterials, rural minor arterials and rural major collectors explain 69.0, 72.7 and 83.7 percent variation in AADT, respectively, by the use of their included X-variable(s).

Using the elasticities obtained from the regression analysis, the forecasting model was developed for each

Table 5.14

Final Regression Equations from Aggregate Analysis (*)

1. Rural Interstate:

AADT = -65569.684 + 0.015365 State Population

R^2 = 0.658 | t = 6.80259

F = 46.275 | e = 4.83314

2. Rural Principal Arterial:

AADT = -27899.830 + 0.339113 County Population + 0.004429 State Population

R^2 = 0.690 | t = 7.82448 | t = 2.22777

F = 40.017 | e = 1.47809 | e = 2.79823

3. Rural Minor Arterial:

AADT = 659.722 + 0.315692 County Household

R^2 = .727 | t = 11.53848

F = 133.136 | e = 0.83377

4. Rural Major Collector:

AADT = -7048.270 + 0.271510 County Population

R^2 = 0.837 | t = 13.41872

F = 180.062 | e = 3.77379

(*) For unit and symbol of each variable, see Table 4.1 of Chapter 4.

category of highway by substituting those elasticities into equation 3.1 (Chapter 3). The models are presented in Table 5.15. These models generally satisfy all the criteria specified earlier. Each of the models is relatively simple, containing not more than two variables. The use of these models is also straightforward. The input values are the present year AADT and the present and future year value (the year for which the traffic forecast is needed) of the predictor variables. The data needed to predict rural traffic volumes with these models are readily available at the county, state levels. The models are easily used by anyone with a hand-held calculator; no large computer system is necessary.

The performance of the models in Table 5.15 were tested using data for those Automatic Traffic Record (ATR) stations not used in building the models. In making these trial "predictions", 1970 data were used as "present year" data. Using the appropriate historical values of the predictor variables, forecasts of AADT for the stations not used in model building based on 1970 AADT were computed and compared with the actual values of AADT. The results of the trial forecasts of the models, shown in Table 5.16, indicate that the models perform satisfactorily. The forecasted errors are reasonably small in most of the cases and speak well for reliability of the models. The larger forecast errors in some cases

Table 5.15

Aggregate Traffic Forecasting Models (*)

1. Rural Interstate:

$$AADT_f = AADT_p \, [1 + 4.83314 \, (\Delta \text{ State Population})]$$

2. Rural Principal Arterial:

$$AADT_f = AADT_p \, [1 + 1.47609 \, (\Delta \text{ County Population} + 2.79623 \, (\Delta \text{ State Population})]$$

3. Rural Minor Arterial:

$$AADT_f = AADT_p \, [1 + 0.86077 \, (\Delta \text{ County Households})]$$

4. Rural Major Collector:

$$AADT_f = AADT_p \, [1 + 0.77379 \, (\Delta \text{ County Population})]$$

(*) (i) For unit and symbol of each variable see Table 4.1 of Chapter 4.

(ii) Δ represents change in predictor variable with respect to its present value in fraction. For example $\Delta X = \dfrac{X_f - X_p}{X_p}$, where X_p and X_f denote present and future values of X.

Table 5.16

Performance of Aggregate Traffic Forecasting Model

(1) Rural Interstate:

Traffic Count Station	Base Year	Year	Forecasted AADT ($AADT_f$)	Actual AADT ($AADT_a$)	Forecast error in percent (*)
5474A	1970	1971	5664	5627	0.66
		1972	5894	6220	-5.24
		1973	6060	6888	-12.02
		1974	6165	6556	-5.96
		1975	6170	6917	-10.80
		1976	6276	7448	-15.74
		1977	6441	7465	-13.72
		1978	6647	7523	-11.64
		1979	6792	7295	-6.90
		1980	6868	6921	-0.64
		1981	6862	6748	1.69
		1982	6827	6745	1.22

* "+" sign indicates overprediction and

"-" sign indicates underprediction.

$$\text{Forecasted error in percent} = \frac{AADT_f - AADT_a}{AADT_a} \times 100.$$

Table 5.16(continued)

(2) Rural Principal Arterial:

Traffic Count Station	Base Year	Year	Forecasted AADT ($AADT_f$)	Actual AADT ($AADT_a$)	Forecast error in percent (*)
173A	1970	1971	10846	10988	-1.29
		1972	11242	11545	-2.62
		1973	11480	12515	-8.27
		1974	11661	11692	-0.27
		1975	11623	11433	1.66
		1976	11685	12396	-5.74
		1977	11917	12872	-7.42
		1978	12335	13065	-5.59
		1979	12599	12391	1.68
		1980	12690	11486	10.48
		1981	12584	11809	6.56
		1982	12442	11607	7.19

* "+" sign indicates overprediction and

"-" sign indicates underprediction.

$$\text{Forecasted error in percent} = \frac{AADT_f - AADT_a}{AADT_a} \times 100.$$

Table 5.16(continued)

(3) Rural Minor Arterial:

Traffic Count Station	Base Year	Year	Forecasted AADT ($AADT_f$)	Actual AADT ($AADT_a$)	$AADT_f - AADT_a$	Forecast error in percent (*)
279A	1970	1971	4825	4848	-23	-0.47
		1972	4955	4946	9	0.18
		1973	5092	4983	109	2.19
		1974	5150	4612	538	11.67
		1975	5169	4644	525	11.30
		1976	5219	4988	231	4.63
		1977	5349	4893	456	9.32
		1978	5471	5225	246	4.71
		1979	5572	5038	534	10.60
		1980	5656	4591	1065	23.19
		1981	5686	4338	1348	31.07
		1982	5772	4419	1353	30.62
319A	1980	1970	1760	1566	194	12.38
		1971	1831	1600	231	14.44
		1972	1861	1652	209	12.65
		1973	1891	2086	-195	-9.35
		1974	1923	1720	203	11.80
		1975	1961	1905	56	2.94
		1976	2004	1947	57	2.93
		1977	2039	2066	-27	-1.31
		1978	2080	2214	-134	-6.05
		1979	2133	2324	-191	-8.22
		1981	2211	2068	143	6.91
		1982	2241	2047	194	9.48
42A	1980	1972	3804	3956	-152	-3.84
		1973	3816	3829	-13	-0.34
		1974	3915	3939	-24	-0.60
		1975	3950	4196	-246	-5.86
		1976	4041	4546	-505	-11.11
		1977	4127	4665	-538	-11.53
		1978	4211	4327	-116	-2.68
		1979	4301	4360	-59	-1.35
		1981	4422	4529	-107	-2.36
		1982	4555	4432	123	2.78

Table 5.16 (continued)

(3) Rural Minor Arterial(Cont'd):

Traffic Count Station	Base Year	Year	Forecasted AADT ($AADT_f$)	Actual AADT ($AADT_a$)	$AADT_f - AADT_a$	Forecast error in percent (*)
100X	1980	1971	8464	8251	213	2.58
		1972	8502	7945	557	7.01
		1973	8602	8402	712	5.07
		1974	8648	8187	461	5.63
		1975	8696	8075	621	7.69
		1976	8784	8611	173	2.01
		1977	8817	8924	-107	1.20
		1978	8880	9454	-574	-6.07
		1979	8961	9389	-428	-4.56
		1981	9005	9022	-17	-0.19
		1982	9041	8837	204	2.31
256A	1970	1971	2652	2738	-86	-3.14
		1972	2729	2710	19	0.70
		1973	2734	2714	19	0.70
		1974	2796	2524	272	10.78
		1975	2840	2709	131	4.84
		1976	2884	2771	113	4.08
		1977	2874	2827	47	1.66
		1978	2957	2940	17	0.58
		1979	2948	2913	35	1.20
		1980	3007	2861	146	5.10
		1981	3045	2925	120	4.10
		1982	3055	2900	155	5.34

* "+" sign indicates overprediction and

"-" sign indicates underprediction.

$$\text{Forecasted error in percent} = \frac{AADT_f - AADT_a}{AADT_a} \times 100.$$

Table 5.16(continued):

(4) Rural Major Collector:

Traffic Count Station	Base Year	Year	Forecasted AADT ($AADT_f$)	Actual AADT ($AADT_a$)	$AADT_f - AADT_a$
7047A	1970	1971	266	257	9
		1972	296	227	69
		1973	292	233	59
		1974	281	226	55
		1975	271	225	46
		1976	271	231	40
		1977	271	204	67
		1978	271	224	47
		1979	241	294	-53
		1980	236	299	-63
		1981	226	288	-62
		1982	205	272	-67
30063A	1980	1979	752	877	-125
		1981	800	824	-24
		1982	793	767	26
54382A	1980	1979	1062	1159	-97
		1981	984	973	11
		1982	1029	878	151
200X	1980	1973	6547	8805	-2258
		1974	6823	8834	-2011
		1975	7155	9002	-1847
		1976	7431	9033	-1602
		1977	8038	9079	-1041
		1978	8535	9457	-922
		1979	8977	9636	-659
		1981	9197	9226	-29
		1982	9308	9004	304

* "+" sign indicates overprediction and

"-" sign indicates underprediction.

$$\text{Forecasted error in percent} = \frac{AADT_f - AADT_a}{AADT_a} \times 100.$$

are due to fewer cases and large variations in response and predictor variables employed in data tables among the stations and counties.

It must be kept in mind that the end use for the forecasted volumes is the design and planning of rural highway projects. These volumes are generally low enough so that larger prediction errors (on the order of 20% to 50%) will not cause a significant change in the design criteria. If more years of data had been available, a better comparison of forecasting models with extrapolations might have been possible. However, this exercise prepares us for another comparison -- aggregate vs. disaggregate models -- to be conducted indirectly later in this chapter.

5.4 Disaggregate Analysis

In this section, each station has been analyzed separately and a separate forecasting model has been developed for each. The criteria for variable selection are the same as that in the aggregate analysis. Performance of the models has been tested with 1983 and 1984 data, which were not used in the development of the models. In disaggregate analysis, the number of observations on which to base each station's model is much smaller than in aggregate analysis, where some of the stations' observations were combined under a highway category. Furthermore, the range in X-variable values is smaller. The key issue here is whether the added consistency in using data from a single station will be enough to offset the reduced amount and range of data values.

No attempt was made to develop disaggregate model for stations 30063A and 54382A under Rural Major Collectors, since only four observations of AADT were available for each of these stations. Also no disaggregate model was developed for stations 313A and 47A. For these two stations, the AADT values were found almost constant over the period of analysis, which was not the case with the predictor variables. Complexity of statistical analysis arises as the number of variables increases and the number of observations decreases. To avoid this complexity, the

variables X_{11} (US Consumer Price Index) and X_{12} (Gross National Product) were dropped from the data tables for Rural Interstates and Rural Principal Arterials. These variables were dropped here because they had failed to survive during the variable selection process in the aggregate analysis. The variable X_3 (Year) has been kept in the data tables to study the residual pattern against X_3.

The analysis starts with the study for scatterplots of AADT (Y) against X's at each station. The scatterplots were done with the help of SPSS [38] to identify any apparent trends of Y with X's. In general, scatterplots of all stations show a linear trend, except for stations 47A, 262A, 279A, 313A and 7047A, which are more scattered. Two representative plots of stations 68A and 7047A are presented in Appendix D. Plots of AADT against Gas Price, as shown in Figures D2 and D13 in Appendix D, were found to be very scattered, which indicates that gas price is less effective to predict AADT than other predictor variables. A slight decrease in AADT from its increasing trend is noticed in the scatterplots at years after 1980. A similar decrease was also observed in some X's (for example, when X_1, X_2 and X_7 are plotted against year).

5.4.1 Multiple Linear Regression Analysis

The same kind of analyses have been done in this section for each station as were done in aggregate analysis for each highway category. The interpretation and selection criteria presented during aggregate analysis are also applicable in disaggregate analysis.

The multiple linear regression analysis starts with the study of the correlation matrix. The SPSS [39] regression program was used to obtain the correlation matrix. Table 5.17 shows the correlation coefficients for the stations under analysis. In general, Table 5.17 shows that the independent variables (X′s) are highly correlated among themselves. The Year (X_3) has low, moderate and high correlation (for example, station 262A: $r = 0.011$, station 134A: $r = 0.429$, station 173A: $r = 0.973$) with AADT (Y). In general, most of the independent or predictor variables (X′s) have high correlation with AADT (Y), except for stations 279A and 262A (Rural Minor Arterials), 47A and 7047A (Rural Major Collectors). But, there is low correlation and, for some stations, negative correlation of X′s with Y (for example, stations: 262A, 279A and 7047A). The reason for this low and/or negative correlation is reflected in Tables A3 and A4 and in the scatter plots of Figures D12 to D17. The AADT (Y) for the above stations remained almost unchanged and, in some cases, decreased during the years 1970 to 1982. However,

Table 5.17

Correlation Matrix (*)
(Disaggregate Analysis)

A. RURAL INTERSTATE

(i) Correlation Coefficients for Station 172A

x1	78657										
x2	44560	.82290									
x3	73803	99084	82725								
x4	66438	.97831	87034	98838							
x5	67997	.98284	.85369	.99477	.99819						
x6	70681	.92713	.71891	.88816	88401	.88759					
x7	85699	.99148	.77891	97463	.94768	.95502	.91079				
x8	84053	98930	.79744	.97405	.94771	.95585	91547	.99299			
x9	74423	99407	.83042	99909	.98787	.99401	.90313	.97787	.98014		
x10	80340	83682	62616	.76280	74739	.74602	.92337	.85716	.84716	.78077	
x13	84256	99057	.75449	.97428	95163	.95889	.92748	.99311	.99044	.97915	.85688
	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x1(

(ii) Correlation Coefficients for Station 3070A

x1	84672										
x2	.37651	75334									
x3	76864	77079	82725								
x4	80809	98524	82747	.97700							
x5	79355	.98499	83598	.98771	.99735						
x6	75002	95242	77558	.93384	.94097	.94449					
x7	84589	99533	.77891	97463	.99280	99144	.96010				
x8	83436	.98462	.79744	97405	98539	.98609	.96625	.99299			
x9	.76997	.97271	83042	99909	.97790	99015	.94580	.97787	.98014		
x10	67961	85512	.62616	76280	.82916	81368	.91267	.85716	84716	.78077	
x13	84571	98650	.75449	97428	.97920	98216	.96615	.99311	.99044	.97915	.85688
	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10

Table 5.17 (continued)

(iii) Correlation Coefficients for Station 5474A

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x1	.81196										
x2	.31843	.76974									
x3	.64578	.96299	.82725								
x4	.93784	.84605	.47167	.70871							
x5	.73939	.98533	.79153	.98799	.80854						
x6	.71325	.93212	.70946	.91203	.75721	.92554					
x7	.76975	.99270	.77891	.97463	.81136	.98827	.95808				
x8	.73512	.98260	.79744	.97405	.79673	.98445	.95365	.99299			
x9	64394	.96346	.83042	.99909	.70907	98728	.92257	97787	.98014		
x10	.67681	.81969	.62616	.76280	.71130	79129	.92698	.85716	.84716	.78077	
x13	.74506	.97859	.75449	.97428	.77562	.98036	.95818	.99311	.99044	.97915	.85688

B. RURAL PRINCIPAL ARTERIAL

(i) Correlation Coefficients for Station 68A

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x1	.94341										
x2	.65503	.80432									
x3	.92558	.98440	.82725								
x4	89925	.98750	.84277	.98830							
x5	90076	.98602	.84200	.99290	.99714						
x6	86713	87340	.65979	81615	.84950	.83673					
x7	.97730	.98367	.77891	.97463	.95611	.95783	.86611				
x8	.97065	.97398	.79744	97405	.75647	.95875	.88650	99299			
x9	.93126	98699	.83042	.99709	99029	.99385	.83826	97787	.98014		
x10	.86743	84730	.62616	.76280	.78023	76733	.91720	85716	84716	.78077	
x13	.97575	.98382	.75449	.97428	.96209	.96320	.88793	.99311	.99044	.97915	.85688

Table 5.17 (continued)

(ii) Correlation Coefficients for Station 134A

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x1	.59559										
x2	.00336	.75939									
x3	.42917	.95580	.82725								
x4	.57527	.92531	.71548	.83329							
x5	.43317	.96653	.83713	.99642	.87136						
x6	.69262	.94862	.61720	.83426	.93153	.86004					
x7	.57625	.99413	.77891	.97463	.90143	.97907	.92488				
x8	.54907	.99039	.79744	.97405	.93149	.98507	.92048	.99299			
x9	.43460	.96223	.83042	.99909	.85083	.99890	.84893	.97787	.98014		
x10	.58558	.88923	.62616	.76280	.86322	.79276	.95600	.85716	.84716	.78077	
x13	.58230	.98759	.75449	.97428	.89747	.98059	.92292	.99311	.99044	.97915	.85688

(iii) Correlation Coefficients for Station 173A

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x1	.97100										
x2	.70032	.82915									
x3	.97254	.98614	.82725								
x4	.11595	.12860	.32846	.21995							
x5	.90149	.91662	.83767	.95852	.48747						
x6	.91330	.91135	.76794	.91460	.30030	.90461					
x7	.98087	.99480	.77891	.97463	.07787	.89069	.91681				
x8	.97320	.99161	.79744	.97405	.14865	.90931	.92632	.99299			
x9	.97409	.98794	.83042	.99909	.23110	.96035	.92769	.97787	.98014		
x10	.82452	.82547	.62616	.76280	-.04537	.66382	.88027	.85716	.84716	.78077	
x13	.98746	.98630	.75449	.97428	.14510	.91005	.93474	.99311	.99044	.97915	.85688

(iv) Correlation Coefficients for Station 254B

	y	x1	x2	x3	x4	x5	x6	x7	x8	x9	x10
x1	.77931										
x2	.49057	.76336									
x3	.79722	.97139	.82725								
x4	.71897	.96153	.79962	.97382							
x5	.77360	.96743	.82604	.99595	.98930						
x6	.82113	.96130	.65566	.91171	.90173	.90653					
x7	.78821	.99879	.77891	.97463	.96234	.97048	.96671				
x8	.78701	.98966	.79744	.97405	.97556	.97685	.96079	.99299			
x9	.80758	.97345	.83042	.99909	.97445	.99563	.92111	.97787	.98014		
x10	.73082	.84427	.62616	.76280	.71979	.74055	.91691	.85716	.84716	.78077	
x13	.83693	.99039	.75449	.97428	.95774	.96993	.96858	.99311	.99044	.97915	.85688

Table 5.17 (continued)

C. RURAL MINOR ARTERIAL

(i) Correlation Coefficients for Station 25A

x1	.83638					
x2	.36205	.75395				
x3	69528	.96122	82725			
x4	.79890	98676	.80572	.96315		
x5	74007	.97990	.83026	.99229	.98847	
x6	.79281	84551	.53144	.74806	.82204	.79053
	y	x1	x2	x3	x4	x5

(ii) Correlation Coefficients for Station 279A

x1	-.14435					
x2	- 67354	.71843				
x3	- 36449	.95979	.82725			
x4	-.23464	.98033	.78397	.94552		
x5	-.34090	.97390	.82893	.99408	97330	
x6	.34805	.84149	.27336	.67956	.78502	.71510
	y	x1	x2	x3	x4	x5

(iii) Correlation Coefficients for Station 301A

x1	.78648					
x2	.37628	.80773				
x3	.72152	.99011	.82725			
x4	81044	.96363	.76825	96921		
x5	.76821	.99430	.80615	.99334	99098	
x6	.80360	.96064	.77978	.94614	.92618	.94400
	y	x1	x2	x3	x4	x5

(iv) Correlation Coefficients for Station 319A

x1	.85658					
x2	.53870	.81982				
x3	.79787	.98915	.82725			
x4	.82157	.98276	.82266	.98158		
x5	.79497	.98799	.83754	.99778	.98909	
x6	.88963	.97611	.73926	.94305	.95418	.94343
	y	x1	x2	x3	x4	x5

Table 5.17 (continued)

(v) Correlation Coefficients for Station 42A

x1	.72487					
x2	.43892	.78383				
x3	.71902	.97272	79713			
x4	70575	.94392	.75351	.98309		
x5	.70145	95990	.78355	.99633	.99384	
x6	.54981	90619	.69489	.90749	.92962	91638
	y	x1	x2	x3	x4	x5

(vi) Correlation Coefficients for Station 100X

x1	.83416					
x2	.44632	.69560				
x3	.76272	.89162	.82297			
x4	-.67455	-.79993	-.83231	-.98141		
x5	81099	.93089	.79869	.99185	-.94930	
x6	.88684	90417	.55097	.73029	-.62910	.78416
	y	x1	x2	x3	x4	x5

(vii) Correlation Coefficients for Station 256A

x1	79504					
x2	47567	.81139				
x3	80099	.96389	82725			
x4	65327	.91358	.70124	.87405		
x5	78159	.98751	.81796	.99082	93046	
x6	81885	.95091	.76557	.94277	83054	.93645
	y	x1	x2	x3	x4	x5

(viii) Correlation Coefficients for Station 262A

x1	10193					
x2	- 34653	.79822				
x3	.01137	.98585	.82725			
x4	.08387	.98902	81765	99083		
x5	03900	.98429	.82977	99755	.99640	
x6	.14587	.95217	.75042	92689	.94005	.93066
	y	x1	x2	x3	x4	x5

Table 5.17 (continued)

D. RURAL MAJOR COLLECTOR

(i) Correlation Coefficients for Station 59A

x1	84968					
x2	.52371	.75334				
x3	.74815	97079	.82725			
x4	.84250	96524	.82747	97700		
x5	.81155	.98467	.83672	.98977	99731	
x6	80040	95242	.77558	93384	94097	.94503
	y	x1	x2	x3	x4	x5

(ii) Correlation Coefficients for Station 200X

x1	79710					
x2	.43441	.64980				
x3	.59079	94177	74156			
x4	.74701	.98249	.74978	.96985		
x5	.64486	95903	.75746	.99567	98718	
x6	86823	81529	42369	.64052	.75357	.68577
	y	x1	x2	x3	x4	x5

(iii) Correlation Coefficients for Station 5420A

x1	.63468					
x2	.38876	.76977				
x3	.58263	.96309	82725			
x4	.60278	.84599	.47167	.70871		
x5	62012	.98538	.79146	.98796	80863	
x6	82502	.93209	.70946	.91203	75721	92576
	y	x1	x2	x3	x4	x5

(iv)) Correlation Coefficients for Station 7047A

x1	.25854					
x2	.63090	.77767				
x3	.39755	.95582	.82725			
x4	-.72214	-.66272	-.80579	-.82225		
x5	-.15553	.86280	.47456	.75134	-.24244	
x6	.49843	.88997	.74353	.91368	-.71662	.73136

(*) For definition of variables, see Table 4.1; Xi represents X_i, where i = 1 to 10 and 13.

the predictor variables (X's) were found to increase over that period. As a result, the X's were less effective in explaining AADT for stations 262A, 279A and 7047A. Thus, if the historical data of AADT for a point or section of highway for which a forecast is desired are available, then the extrapolation of the plot of AADT against time at future year will detect any unreasonable value of future AADT computed from the forecasting model(s). If the change in AADT is not significant over a period of time, then it will be reasonable to assume that the future value of AADT will not be changed significantly. In that case, using predictor variables that increase significantly over a period of time will overestimate the future year AADT. Then, simple extrapolation of the plot of AADT against time will provide better results. In spite of reduced effectiveness of individual X's to predict Y for stations 262A and 279A, further analyses have been carried out for these stations because combination of X's may provide better results for some stations.

It was noticed during the aggregate analysis that the case A stepwise regression with default parameters, as defined in section 5.3.1.2, has little control over variable selection and almost all the variables were entered in that case (see Table 5.4). As a result, only case B and case C stepwise regressions, defined in section 5.3.1.2, were carried out for the stations under this

disaggregate analysis with the help of the SPSS package [39]. But, the case A stepwise regression was done only for those stations for which no variables remained in the equation after the case B and case C stepwise regressions. The summary of the stepwise regression analysis is shown in Table 5.18.

The C_p-statistic, R^2, etc. in the all possible regressions were calculated with the help of a program "DRRSQU" [42]. Some of the selected values of C_p and R^2 are presented in Table 5.19. Variable X_3 (Year) and its combinations with other X-variables were not presented in Table 5.19. The variable X_3 was kept only for graphic residual analysis. Moreover, year as a predictor variable is not suitable because it will always increase, which is not true for AADT (Y). The values of the other X's in the data tables could be increase or decrease, as AADT does over the years. Moreover, the effect of X_3 is reflected in some other X's.

5.4.2 Preliminary Screening of Candidate Variables

The screening of the variables has not been done solely on the basis of statistical analysis. Subjective judgments regarding the questions in Table 5.6 have always been included in the selection process, as was done during the aggregate analysis. Introduction of subjective judgment into the forecasting process is one of the

Table 5.18

Stepwise Regression Summary (Disaggregate Analysis)

Highway Category	ATR Station	Case (*)	Step	Variable subscript En-tered	Variable subscript Re-moved	F value	Signifi-cance level	last step b-coeff. (**)	R Squared	Overall F	Overall signifi-cance
			1	7		30.421	0.0	.0168	0.734	30.421	0.0
RURAL	172A	B & C	2	1		76.767	0.0	-2.2489	0.969	158.364	0.0
			3	13		9.275	0.014	6.9492	0.985	196.032	0.0
			Constant	term				-13999.34			
			1	1		27.861	0.0		0.717	27.861	0.0
			2	2		12.626	0.005	-194.2593	0.875	34.966	0.0
INTERSTATE	3070A	B & C	3	4		6.310	0.033	.7796	0.926	37.792	0.0
			4		1	0.331	0.579		0.924	60.575	0.0
			Constant	term				-6035.972			
	5474A	B & C	1	4		80.320	0.0	1.1889	0.880	80.320	0.0
			Constant	term				-34973.79			
			1	7		234.112	0.0	.0019	0.955	234.112	0.0
RURAL	68A	B & C	2	2		17.697	0.002	-25.7583	0.984	303.583	0.0
			Constant	term				747.0824			
			1	6		10.143	0.009		0.480	10.143	0.009
			2	2		12.651	0.005	-87.6167	0.770	16.768	0.001
PRINCIPAL			3	7		4.989	0.052	.0049	0.852	17.301	0.0
	134A	B & C	4		6	0.948	0.356		0.837	25.611	0.0
			5	9		9.069	0.015	-.0104	0.919	33.874	0.0
			Constant	term				15267.25			
ARTERIAL	173A	B & C	1	13		430.373	0.0	3.5697	0.975	430.373	0.0
			Constant	term				-3992.541			
			1	13		25.723	0.0	8.8341	0.700	25.723	0.0
	254B	B & C	2	7		8.150	0.017	-.0051	0.835	25.296	
			Constant	term				-9374.137			
			1	1		25.609	0.0	.1954	0.700	25.609	0.0
RURAL		B	2	2		12.528	0.005	(-16.711)	0.867	32.489	0.0
			3	3		3.710	0.086	(-67.303)	0.906	28.766	0.0
			Constant	term				(132329.)			
	25A										
			1	1		25.609	0.0	.1286	0.700	25.609	0.0
		C	2	2		12.528	0.005	-24.9935	0.867	32.489	0.0
MINOR			Constant	term				1492.571			
			1	2		9.134	0.012		0.454	9.133	0.012
			2	6		12.740	0.005	.0392	0.760	15.811	0.001
		B	3	5		3.963	0.078	-.0949	0.833	14.985	0.001
			4		2	0.009	0.928		0.833	24.945	0.0
ARTERIAL	279A		Constant	term				6534.579			
			1	2		9.134	0.012	-28.5409	0.454	9.133	0.012
		C	2	6		12.740	0.005	.0186	0.760	15.811	0.001
			Constant	term				4892.497			

Table 5.18 (continued)

Highway Category	ATR Station	Case (*)	Step	Variable subscript		F value	Significance level	last step b-coeff. (**)	R Squared	Overall F	Overall significance
				Entered	Removed						
RURAL	301A	B	1	4		21.052	.001	.2561	.657	21.052	.001
			2	2		7.569	.020	-16.1671	.865	20.626	0.0
			3	6		4.766	.057	.1955	.872	20.519	0.0
			4	3		4.253	.073	(-64.965)	.917	22.014	0.0
			Constant term					(125786.)			
		C	1	4		21.052	.001	.2362	.657	21.052	.001
			2	2		7.569	.029	-17.9281	.865	20.626	0.0
			Constant term					(-988.27)			
	319A	B & C	1	6		41.743	0.0	.0899	.791	41.743	0.0
			Constant term					761.2511			
MINOR ARTERIAL	42A	B & C	1	1		9.965	.012	.0279	.525	9.965	.012
			Constant term					2147.178			
	100A	B & C	1	6		36.837	0.0	.0242	.766	36.837	0.0
			Constant term					4105.917			
	256A	B & C	1	6		22.386	.001	.1642	.671	22.386	.001
			Constant term					2414.798			
	262A	A	1	2		1.501	.246	-11.4293	.120	1.501	.246
			2	4		8.601	.015	(0.2365)	.527	5.570	.024
			3	3		4.006	.076	(-71.635)	.673	6.164	.015
			4	1		.849	.384	(0.0436)	.784	4.756	.029
			Constant term					(138352.)			
		B & C	(NO VARIABLES REMAIN IN THE EQUATION)								

Table 5.18 (continued)

Highway Category	ATR Station	Case (*)	Step	Variable subscript Entered	Variable subscript Removed	F Value	Signifi-cance level	last step b-coeff. (**)	R Squared	Overall F	Overall signifi-cance
RURAL MAJOR COLLECTOR	59A	B	1	1		28.563	0.0		.722	28.563	0.0
			2	3		5.818	.037	-120.1901	.824	23.445	0.0
			3	4		3.457	.096	.2672	.873	20.622	0.0
			4		1	2.882	.128		.863	25.024	0.0
			5	2		5.983	.033	-19.7815	.899	26.829	0.0
			Constant term					231941.55			
		C	1	1		28.563	0.0	.1213	.722	28.563	0.0
			2	3		5.818	.037	-109.9092	.824	23.445	0.0
			Constant term					217571.57			
	200A	B & C	1	6		24.497	.001	.03147	.754	24.497	.001
			Constant term					6614.1961			
	5420A	B & C	1	6		23.447	.001	.2462	.681	23.447	.001
			2	3		11.596	.007	-63.3964	.853	26.814	0.0
			Constant term					124756.32			
	7847A	B	1	4		11.998	.005	-.0574	.521	11.998	.005
			2	3		3.386	.099	-16.1501	.640	8.903	.006
			3	6		10.669	.010	.4620	.835	15.201	.001
			4	2		7.061	.023	2.3385	.917	22.095	0.0
			Constant term					33254.73			
		C	1	4		11.988	.005	-.0400	.521	11.998	.005
			Constant term					1122.064			

(*) Case A: (All Default Parameters)

1. Max. No. of Steps = 2 * No. of Independent Variables
2. FIN = .01; FOUT = .005
3. Tolerance level = .001

Case B:

1. Max. No. of Steps = 2 * No. of Independent Variables (Default)
2. FIN/FOUT = F(.10 1,n-p);where, FIN > FOUT, n = No. of cases, p = No. of Expected Parameter in Equation.
3. Tolerance level = .01

Case C:

SAME AS CASE B Except FIN/FOUT = F(.05,1, n-p)

(**) 95% Confidence Interval of the b-coefficient in ( ) includes zero.

Table 5.19

Selected C_p & R-Squared in All Possible Regression
(Disaggregate Analysis)

Highway Category	ATR Station	Subscripts of Variables in Equation	C_p Values in same order	R-Squared Values in same order	P
RURAL	172A	7, 13, 8,	13476,14722,14895,	.734, .710, .706,	2
		10, 1, 9	17995,19354,22645	.645,.619, .554	
		1 7,. 5 7, 4 7,	1547, 2407, 2597,	.969, .952, .949,	3
		7 9, 5 13, 4 13	3272, 4402, 4565,	.935, .913, .910,	
		5 8	5937	.883	
		1 7 13, 1 2 7, 1 7 8,	760, 1314, 1335,	.985, .974, .974,	4
		1 6 7, 1 5 7, 1 7 10,	1454,1456, 1467,	.971,.971, .971,	
		1 4 7, 1 7 9, 5 6 7,	1497, 1511, 1772,	.970, .970, .965,	
		5 7 10, 2 5 7, 4 5 9	1898, 1932, 4688	.963, .962, .908	
		....	......	......	...
		Subscripts of all variables	12	1.000	12
INTERSTATE	3070A	1, 7, 13,	34.60,34.82,34.86,	.717, .716, .715,	2
		8, 4, 5,	37.80,44.45,48.04,	.698, .653, .630,	
		9	53.71	.593	
		2 8, 2 4, 2 7,	4.50, 4.75, 5.59,	.925, .924, .918,	3
		2 5, 1 2, 2 13,	7.94,12.27,12.40,	.903, .875, .874,	
		5 7, 2 9	18.44,21.42	.835, .816	
		2 6 8, 2 6 7, 2 8 10,	0.18, 4.00, 4.04,	.966, .942, .941,	4
		5 7 10, 2 4 8, 2 7 10,	4.20, 4.21, 4.25,	.940, .940, .940,	
		1 2 7, 2 4 10, 2 5 8	5.31, 5.66, 5.74	.933, .931, .93	
		....	......	......	...
		Subscripts of all variables	12	.994	12
	5474A	4, 1, 7,	72.0,220.2,265.1,	.880, .659, .593,	2
		11, 5, 8	290.2,295.9,300.1	.555, .547, .54	
		1 9, 2 4, 1 2,	42.7, 60.7, 67.0,	.926, .899, .890,	3
		1 4, 4 5, 4 13	73.2, 73.3, 73.5	.881, .881, .88	
		1 2 9, 1 2 4, 7 9 10,	3.32,25.07,31.29,	.988, .955, .946,	4
		1 9 13, 1 5 9, 1 4 9,	32.88,32.98,33.36	.944, .944, .943	
		....	......	......	...
		Subscripts of all variables	12	.999	12

Table 5.19 (continued)

Highway Category	ATR Station	Subscripts of Variables in Equation	C_p Values in same order	R-Squared Values in same order	P
RURAL PRINCIPAL ARTERIAL	68A	7, 13, 8,	55.0, 59.4, 73.6,	.955, .952, .942,	2
		1, 9, 5,	148.0,180.5,260.2,	.890, .867, .811,	
		4, 10, 6	264.1,344.3,345.1	.809, .752, .752	
		2 7, 2 8, 5 13,	16.12,20.04,31.20,	.984, .981, .973,	3
		4 13, 5 7, 4 7,	31.43,35.47,36.50,	.973, .970, .970,	
		7 9, 1 8	37.61,75.45	.969, .942	
		2 7 8, 1 2 7, 2 5 7,	13.04,15.23,15.98,	.987, .986, .985,	4
		2 4 7, 2 7 9, 1 2 8	16.37,16.59,20.56	.985, .985, .982	
		....	......	......	...
		Subscripts of all variables	12	.999	12
	134A	2 7, 5 13, 1 2,	36.57,37.52,38.14,	.837, .833, .831,	3
		2 8, 9 13, 2 13,	41.01,50.49,51.59,	.820, .784, .780,	
		5 8, 1 8	74.88,141.9	.693, .442	
		2 7 9, 2 5 7, 2 5 8,	16.70,17.36,17.56,	.919, .916, .915,	4
		4 5 9, 2 8 9, 5 8 9	24.08,26.30,26.75	.891, .883, .881	
		....	......	......	...
		Subscripts of all variables	12	.996	12
	173A	13, 7, 9,	27.60,46.64,66.11,	.975, .962, .949,	2
		8, 1, 5	68.67,74.94,266.1	.947, .943, .813	
		2 9, 1 2, 9 13,	12.32,25.34,27.75,	.987, .978, .976,	3
		8 13, 4 13, 1 13,	27.81,28.48,29.14,	.976, .976, .975,	
		5 13, 1 8	29.53,67.54	.975, .949	
		2 9 10, 2 7 9, 1 2 9,	2.38, 4.08, 5.83,	.995, .994, .993,	4
		2 9 13, 2 5 9, 2 4 9	5.89, 6.28,8.07	.993, .992, .991	
		....	......	......	...
		Subscripts of all variables	12	.999	12
	254B	9, 7, 8,	16.04,18.26,18.40,	.652, .621, .619,	2
		1, 5, 4	19.27,19.90,25.77	.607, .598, .517	
		7 13, 1 13, 8 13,	4.88, 5.31, 7.92,	.835, .829, .793,	3
		4 13, 2 9, 4 9,	8.63,10.52,11.44,	.783, .757, .744,	
		5 13, 1 8	12.79,20.39	.725, .619	
		2 5 9, 4 7 13, 2 4 9,	4.64, 4.67, 4.76,	.866, .866, .864,	4
		1 4 13, 1 8 13, 1 2 13	4.98, 5.54, 5.62	.861, .854, .852	
		....	......	......	...
		Subscripts of all variables	12	.986	12

Table 5.19 (continued)

Highway Category	ATR Station	Subscripts of Variables in Equation	C_P Values in same order	R-Squared Values in same order	P
R U R A L	25A	1, 4, 6, 5	24.56, 31.43, 32.52, 41.55	.700, .638, .629, .546	2
		1 2, 2 4, 1 5, 2 5, 4 5, 1 4	7.91, 8.16, 8.83, 20.63, 21.43, 23.62	.867, .860, .858, .753, .746, .726	3
		1 2 5, 2 4 5, 1 4 5, 1 2 4, 1 2 6, 2 4 6	7.11, 7.30, 8.72, 8.98, 9.74, 9.77	.892, .890, .877, .875, .868, .868,	4
		.... Subscripts of all variables	 7	...... .946	... 7
	279A	5 6, 1 5, 4 6, 1 6, 2 6, 1 2, 2 4, 2 5	0.33, 1.79, 2.07, 2.82, 3.54 6.52, 7.17, 10.34	.833, .800, .793, .776, .760, .692, .677, .665	3
		4 5 6, 2 5 6, 1 2 5, 1 4 5, 1 4 6, 1 2 6	2.19, 2.32, 2.55, 3.16, 3.73, 4.27	.836, .833, .828, .814, .801, .769	4
		.... Subscripts of all variables	 7	...... .863	... 7
M I N O R	301A	4, 6, 1, 5	20.16, 21.10, 23.41, 25.83	.657, .646, .619 .59	2
		1 2, 2 6, 2 4, 2 5, 4 5, 1 4	9.01, 9.51, 9.58, 13.49, 16.39, 21.86	.812, .806, .805, .759, .725, .66	3
		2 4 6, 1 2 6, 2 5 6, 4 5 6, 1 2 5, 1 2 4	5.84, 8.17, 8.91, 9.58, 10.12, 10.18	.872, .845, .836, .828, .823, .821	4
		.... Subscripts of all variables	 7	...... .929	... 7
A R T E R I A L	319A	6, 1, 4, 5	5.83, 9.93, 14.11 17.17	.791, .734, .675, .632	2
		1 5, 2 6, 1 2, 5 6, 4 6, 1 6, 1 4, 2 4	4.09, 5.61, 6.13, 6.56, 7.23, 7.62, 11.88, 11.97	.844, .823, .815, .809, .800, .794, .746, .733	3
		1 2 5, 1 4 5, 1 5 6, 1 2 6, 2 5 6, 1 2 4	3.48, 4.87, 5.92, 7.06, 7.55, 7.84	.861, .861, .846, .830, .824, .819	4
		.... Subscripts of all variables	 7	...... .916	... 7

Table 5.19 (continued)

Highway Category	ATR Station	Subscripts of Variables in Equation	C_p Values in same order	R-Squared Values in same order	P
RURAL	42A	1, 4, 5	12.73, 13.87, 14.12	.525, .498, .492	2
		1 6, 4 6, 1 2, 1 4, 1 5, 2 4	12.06, 12.41, 12.93, 14.55, 14.71, 15.64	.590, .581, .569, .538, .526, .518	3
		....	......	......	...
		Subscripts of all variables	7	.984	7
MINOR ARTERIAL	108A	6, 1, 5	6.91, 13.24, 15.9	.786, .695, .658	2
		4 6, 1 6, 2 6, 2 5, 4 5, 1 2	7.34, 8.51, 8.73, 10.08, 11.48, 12.81	.809, .792, .789, .778, .758, .731	3
		....	......	......	...
		Subscripts of all variables	7	.926	7
	256A	6, 1, 5	4.26, 5.81, 6.66	.671, .632, .611	2
		2 6, 1 2, 2 5, 1 6, 4 6, 1 4	4.04, 4.43, 5.48, 6.15, 6.17, 6.51	.726, .716, .692, .673, .673, .664	3
		....	......	......	...
		Subscripts of all variables	7	.851	7
	262A	2 4, 1 2, 2 6	13.13, 13.63, 14.39	.527, .515, .497	3
		2 4 5, 2 4 6, 1 2 6, 1 2 4, 1 2 5, 2 5 6	10.88, 14.48, 14.93, 15.06, 15.49, 15.65	.629, .544, .538, .529, .518, .515	4
		....	......	......	...
		Subscripts of all variables	7	.859	7
RURAL MAJOR COLLECTOR	7047A	4	14.79	.554	2
		5 6, 1 4, 4 5, 2 5, 1 5, 2 4	4.87, 6.25, 7.85, 14.89, 16.23, 16.27,	.814, .778, .737, .593, .566, .565	3
		2 5 6, 1 4 6, 4 5 6, 1 5 6, 1 4 5, 1 2 4	1.36, 3.26, 3.42, 4.38, 6.93, 7.35	.911, .872, .869, .851, .797, .766	4
		....	......	......	...
		Subscripts of all variables	7	.918	7

Table 5.19 (continued)

Highway Category	ATR Station	Subscripts of Variables in Equation	C_p Values in same order	R-Squared Values in same order	P
RURAL	59A	1, 2, 4	21.16, 22.40, 29.96	.722, .710, .641	2
		4 5, 2 4, 1 2, 1 5, 2 5, 1 4	7.90, 14.13, 19.76, 26.91, 21.30, 23.85	.863, .805, .753, .743, .739, .723	3
		2 4 5, 1 4 5, 4 5 6, 1 2 4, 2 4 6, 1 2 6	5.20, 6.30, 8.63, 14.81, 16.10, 21.7	.906, .896, .868, .817, .806, .754	4
		....	......	......	...
		Subscripts of all variables	7	.945	7
MAJOR	208A	6, 1, 4	35.85, 55.93, 69.13,	.754, .635, .558	2
		4 5, 1 5, 1 6, 4 6, 1 2, 2 4	13.95, 27.69, 33.81, 34.46, 55.93, 64.99	.894, .814, .776, .774, .647, .594	3
		4 5 6, 1 4 5, 2 4 5, 1 2 5, 1 5 6, 1 4 6	10.93, 12.78, 14.38, 23.16, 25.33, 35.64	.924, .910, .904, .852, .839, .779	4
		....	......	......	...
		Subscripts of all variables	7	.962	7
COLLECTOR	5420A	6	6.24	.681	2
		5 6, 1 6, 2 6, 4 6	1.35, 1.67, 4.52, 8.18	.825, .818, .758, .682	3
		1 4 6, 4 5 6, 1 2 6, 2 5 6, 1 5 6, 2 4 6	1.13, 2.32, 2.82, 2.87, 3.12, 6.26	.872, .847, .836, .835, .830, .764	4
		....	......	......	...
		Subscripts of all variables	7	.874	7

suggestions made by Armstrong in his critique of common practice [4,5]. The first four goals in Table 5.7 were considered in the preliminary choices of candidates for the final selections. As a general rule [15], the third goal (i.e. number of X-variables in model) does not support more than 2 variables in the equation (See footnote of Table 5.7). In these preliminary choices, the third goal was relaxed for some stations in order to satisfy other goals in Table 5.7. In preliminary choices, similar kinds of diagnoses, as were done in the case of the aggregate analysis in sections 5.3.2.1 to 5.3.2.4, were carried out for the stations under investigation. The results of the preliminary screening process are shown in Table 5.20. The statistical results on R^2 value, C_P-criterion, stepwise regression, and correlation coefficient were not considered alone in making the preliminary choices. Subjective judgments regarding the questions in Table 5.6 were also involved in these preliminary choices.

5.4.3 Final Selection of Variables

The final model selection for each station was made from the preliminary choices of Table 5.20 by examining the goals of the analysis in Table 5.7. The goals 1 to 4 in Table 5.7 were considered during the preliminary screening process. The signs of regression coefficients

Table 5.20

Preliminary Choices of Disaggregate Analysis

Highway Category	ATR Station	Choice Numbers	Subscripts of Variables in same order as Choice Numbers
RURAL INTERSTATE	172A	1, 2, 3, 4, 5	7, 8, 1 7, 4 7, 1 7 13
	3070A	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1, 4, 7, 8, 2 4, 2 7, 2 8, 2 6 7, 2 6 8, 2 8 10
	5474A	1, 2, 3, 4	4, 1 9, 2 4, 1 2 9
RURAL PRINCIPAL ARTERIAL	68A	1, 2, 3	7, 2 7, 2 8
	134A	1, 2, 3, 4	1 2, 2 7, 2 8, 2 7 9
	173A	1, 2, 3, 4, 5	7, 8, 1 2, 4 13, 2 7 9
	254B	1, 2, 3, 4, 5, 6	1, 4, 7, 8, 7 13, 8 13
RURAL MINOR ARTERIAL	25A	1, 2, 3, 4	1, 4, 1 2, 2 4
	279A	1, 2	1 6, 2 6
	301A	1, 2, 3, 4	1, 4, 1 2, 2 4
	319A	1, 2, 3, 4, 5	1, 4, 6, 1 2, 2 4
	42A	1, 2, 3, 4	1, 4, 1 2, 2 4
	100X	1, 2, 3, 4, 5	1, 6, 1 2, 2 6, 4 6
	256A	1, 2, 3, 4	1, 6, 1 2, 2 6
	262A	1, 2	1 2, 2 4
RURAL MAJOR COLLECTOR	59A	1, 2, 3, 4, 5	1, 4, 1 2, 1 4, 2 4
	200X	1, 2, 3, 4, 5	1, 4, 6, 1 2, 2 4
	5420A	1, 2, 3	6, 1 6, 2 6
	7047A	1, 2	4, 1 4

were checked through the regression on preliminary choices of Table 5.20. Final selection for each station was determined by examining all the criteria of Table 5.7 except the residual analysis and hypothesis testing concerning b-coefficients. Graphic residual analysis and tests concerning regression coefficients were carried out on the final selection before transforming into model according equation 3.1 of Chapter 3. Residuals plots were done to check whether some key independent or predictor variables could provide additional predictive power to the models developed. The tests concerning regression coefficients were done to confirm that the variables in the models are statistically significant.

5.4.3.1 Regression on Preliminary Choices and Final Selection

Regression on the preliminary choices was carried out with the help of the SPSS package [39]. The summary of that analysis is shown in Table 5.21 for all the stations under analysis. The magnitudes of b-coefficients and their inconsistency with respect to sign are shown in Table 5.21, together with R^2, overall F value, and significance of the choices. Table 5.21 also shows the variables for which the 95 percent confidence interval of the b-coefficients includes zero. To find the final selection, in which the 95 percent confidence interval does not include zero was preferred. The diagnoses -- carried out on the choices for the stations in Table 5.21

Table 5.21

Multiple Linear Regression Summary
on Preliminary Choices
(Disaggregate Analysis)

Highway Category	ATR Station	Choice Number	Variable Subscript (*)	b-coefficient in same order	Inconsistencies in b's (**)	R Squared	Overall F	Overall Significance
RURAL INTERSTATE	172A	1	7	0.0038	---	0.734	30.421	0.0
		2	8	0.0171	---	0.706	26.478	0.0
		3	1,7	-2.0297, 0.0201	-b1	0.969	158.365	0.0
		4	4,7	-1.9864, 0.0099	-b4	0.949	92.514	0.0
		5	1,7,13	-2.2489, 0.0168, 6.9492	-b1	0.985	196.032	0.0
	3070A	1	1	0.2352	---	0.717	27.861	0.0
		2	4	0.4000	---	0.653	20.702	0.0
		3	7	0.0030	---	0.716	27.668	0.0
		4	8	0.0137	---	0.696	25.204	0.0
		5	2,4	-194.259, 0.7796	---	0.924	60.575	0.0
		6	2,7	-150.500, 0.0050	---	0.918	56.154	0.0
		7	2,8	-166.311, 0.0241	---	0.925	61.951	0.0
		8	2,(6),7	-142.315, -0.5107, 0.0068	-b6	0.942	48.357	0.0
		9	2,6,8	-164.021, -0.7264, 0.0364	-b6	0.966	82.232	0.0
		10	2,8,(10)	-173.163, 0.0278, -0.0032	-b10	0.941	48.123	0.0
	5474A	1	4	1.1889	---	0.880	80.320	0.0
		2	1,9	0.6558, -0.0108	-b9	0.926	62.672	0.0
		3	(2),4	-14.2176, 1.2842	---	0.899	44.650	0.0
		4	1,2,9	0.6086, -40.5611, -0.0076	-b9	0.988	239.613	0.0
RURAL PRINCIPAL ARTERIAL	68A	1	7	0.0016	---	0.955	234.112	0.0
		2	2,7	-25.7583, 0.0019	---	0.984	303.583	0.0
		3	2,8	-31.1802, 0.0092	---	0.981	258.892	0.0
	134A	1	1,2	0.0443, -109.080	---	0.831	24.547	0.0
		2	2,7	-116.511, 0.0026	---	0.837	25.611	0.0
		3	2,8	-122.754, 0.0121	---	0.820	22.778	0.0
		4	2,7,9	-87.6167, 0.0049, -0.0104	-b9	0.919	33.874	0.0
	173A	1	7	0.0026	---	0.962	279.343	0.0
		2	8	0.0119	---	0.947	196.993	0.0
		3	1,2	0.4903, -52.2181	---	0.978	222.058	0.0
		4	(4),13	-0.0592, 3.5844	-b4	0.976	201.959	0.0
		5	2,7,9	-47.5616, 0.0011, 0.0080	---	0.994	482.024	0.0
	254B	1	1	0.1495	---	0.607	17.013	0.0
		2	4	0.3870	---	0.516	11.771	0.006
		3	7	0.0013	---	0.621	18.045	0.001
		4	8	0.0059	---	0.619	17.901	0.001
		5	7,13	-0.0051, 8.8341	-b7	0.835	25.296	0.0
		6	(8),13	-0.0166, 6.7564	-b8	0.793	19.127	0.0

Table 5.21 (continued)

Highway Category	ATR Station	Choice Number	Variable Subscript (*)	b-coefficient in same order	Inconsistencies in b's (**)	R Squared	Overall F	Overall Significance
RURAL	25A	1	1	.0624	---	.700	25.609	0.0
		2	4	.1695	---	.638	19.407	0.0
		3	1, 2	0.1286, -24.9935	---	.867	32.489	0.0
		4	2, 4	-32.2458, .3067	---	.864	31.858	0.0
	279A	1	1, 6	-0.0337, .0521	-b1	.776	17.331	.001
		2	2, 6	-28.5409, .0186	---	.760	15.811	.001
	301A	1	1	.0806	---	.619	17.838	.001
		2	4	.1505	---	.657	21.052	.001
		3	1, 2	0.1423, -22.2131	---	.812	21.530	0.0
		4	2, 4	-17.9201, .2363	---	.805	20.626	0.0
MINOR	319A	1	1	.0471	---	.734	30.311	0.0
		2	4	.1688	---	.675	22.844	.001
		3	6	.0900	---	.791	41.743	0.0
		4	1, (2)	0.0695, -16.3968	---	.815	22.069	0.0
		5	(2), 4	-13.9513, .2405	---	.733	13.740	.001
	42A	1	1	.0279	---	.525	9.965	.012
		2	4	.0503	---	.498	8.931	.015
		3	1, (2)	0.0361, -12.3481	---	.569	5.276	.035
		4	(2), 4	-7.9152, .0656	---	.518	4.299	.054
ARTERIAL	100X	1	1	.0168	---	.696	22.876	.001
		2	6	.0242	---	.786	36.837	0.0
		3	1, (2)	0.0205, -17.2954	---	.731	12.202	.003
		4	(2), 6	-4.0498, .0251	---	.789	16.833	.001
		5	(4), 6	-0.0081, .0209	-b4	.809	19.061	.001
	256A	1	1	.0619	---	.632	18.899	.001
		2	6	.1642	---	.671	22.386	.001
		3	1, (2)	0.1233, -8.4128	---	.716	12.612	.002
		4	(2), 6	-6.1982, .2203	---	.726	13.232	.002
	262A	1	1, 2	0.0356, -12.9828	---	.515	5.309	.027
		2	2, 4	-13.7683, .0672	---	.526	5.570	.024

Table 5.21 (continued)

Highway Category	ATR Station	Choice Number	Variable Subscript (*)	b-coefficient in same order	Inconsistencies in b's (**)	R Squared	Overall F	Overall Significance
RURAL MAJOR COLLECTOR	59A	1	1	.0481	---	.722	28.563	0.0
		2	4	.0850	---	.709	26.906	0.0
		3	1,(2)	0.0596, -11.4951	---	.753	15.267	.001
		4	(1),(4)	0.0379, .0165	---	.723	13.047	.002
		5	(2),4	-23.4952, .1309	---	.805	20.670	0.0
	200X	1	1	.0502	---	.635	13.940	0.0
		2	4	.1080	---	.558	10.101	.013
		3	6	.0815	---	.754	24.497	.001
		4	1,(2)	0.0571, -5.9598	---	.647	6.428	.026
		5	(2),4	-11.8313, .1391	---	.594	5.123	.043
	5420A	1	6	.1163	---	.681	23.447	.001
		2	1,6	-0.0917, .2509	-b1	.818	22.501	0.0
		3	(2),6	-12.8709, .1559	---	.758	15.700	.001
	7047A	1	4	-.0433	---	.521	11.988	.005
		2	(1),4	-0.0091, -.0590	-b1	.608	7.749	.009

(*) 95% confidence interval of the b-coefficent(s) for the variable(s) enclosed in first bracket includes zero.

(**) '+' and '-' sign with b-coefficient(s) are inconsistent with the expected result.

to find the final selection for each station -- were similar to those done earlier for aggregate analysis in Section 5.3.3.1. The final selections from this disaggregate analysis are shown in Table 5.22. In general, not more than two X-variables were taken for the final selection.

5.4.3.2 Graphic Residual Analysis on Final Selections

The stations 3070A, 68A, 301A, 7047A -- one from each of the four highway categories -- were picked through random sampling. The residual plots of these representative stations, shown in Appendix E, were generated by the BMDP package [47]. The plots of other stations were found similar to the plots of these representative stations. The residual plots against the predicted AADT, the final selected predictor variable(s), and the "year" are presented in Figures E1.1 to E1.4, E2.1.1 to E2.4.1, and E4.1 to E4.4, respectively, in Appendix E.

The normal probability plots of residuals are given in Figures E3.1 to E3.4 in Appendix E. These plots appear reasonably close to straight lines and indicate that error terms are approximately normally distributed. The random pattern of plots of residuals against the fitted response variable and predictor variables (Figures E1.1 to E2.4.1) indicate no ground for suspecting the appropriateness of

Table 5.22

Final Selection of Disaggregate Analysis

Highway Category	ATR Station	Variable Subscript	b-coefficient in same order	Inconsistencies in b's (*)	R Squared	Overall F	Overall Significance
RURAL INTERSTATE	172A	8	.0171	---	.706	26.478	0.0
	3070A	2, 6	-166.311, .0241	---	.925	61.951	0.0
	5474A	4	1.1889	---	.880	80.320	0.0
RURAL PRINCIPAL ARTERIAL	68A	7	.0016	---	.955	234.112	0.0
	134A	2, 7	-116.511, .0026	---	.837	25.611	0.0
	173A	1, 2	0.4903, -52.2181	---	.978	222.056	0.0
	254B	1	.1495	---	.607	17.013	0.0
RURAL MINOR ARTERIAL	25A	1, 2	0.1266, -24.9935	---	.867	32.459	0.0
	279A	2, 6	-28.5409, .0186	---	.760	15.811	.001
	301A	1, 2	0.1423, -22.2131	---	.812	21.530	0.0
	319A	1	.0471	---	.734	30.311	0.0
	42A	1	.0279	---	.525	9.965	.012
	100X	1	.0168	---	.696	22.876	.001
	256A	1	.0819	---	.632	18.899	.001
	262A	1, 2	0.0356, -12.9826	---	.515	5.309	.027
RURAL MAJOR COLLECTOR	59A	1	.0481	---	.722	28.563	0.0
	200X	1	.0503	---	.635	13.940	0.0
	5420A	6	.1163	---	.681	23.447	.001
	7047A	4	-.0433	---	.521	11.988	.005

(*) '+' and '-'sign with b-coefficient(s) are inconsistent with the expected result.

the linearity of the regression function or constancy of the error variance.

Residual plots were also generated against variables not included in the model to check whether some key independent or predictor variables had been excluded from any model. One such variable is the Year (X_3), which was not included in any model. The plots of residuals against X_3, shown in Figures E4.1 to E4.4, do not indicate any correlation between error terms over Year, since the residuals are random around the zero line.

5.4.3.3 Testing Hypothesis Concerning Regression Coefficients

The same overall F-test and partial F-test used for the aggregate analysis of each highway category have been applied to each station separately. The results of these two F-tests are shown in Tables 5.23 and 5.24. The partial F-test for one variable is the same as that of the overall F-test. The overall F-test results of Table 5.23 at α-levels of 0.05 and 0.10 show that the regression relationships between the predictor variable(s) and the response variable exist and cannot be rejected at an α-level of as low as 0.05. The partial F-test results for those stations with more than one variable in the regression equations are shown in Table 5.24 at α-levels of 0.05 and 0.10. The results show that the variable(s) in reduced models have significant influence (i.e., cannot

Table 5.23

Overall F-tests for Disaggregate Analysis

Highway Category	ATR Station	Variable Subscripts for Full Model	df_R , df_F (*)	F^*	α	Is H_a true for α = .05?	Is H_a true for α = .10?
Rural Interstate	172A	8	1, 11	26.478	<.001	Yes	Yes
	3070A	2, 6	2, 10	61.951	<.001	Yes	Yes
	5474A	4	1, 11	80.320	<.001	Yes	Yes
Rural Principal Arterial	68A	7	1, 11	234.112	<.001	Yes	Yes
	134A	2, 7	2, 10	25.611	<.001	Yes	Yes
	173A	1, 2	2, 10	222.058	<.001	Yes	Yes
	254B	1	1, 11	17.013	.001-.005	Yes	Yes
Rural Minor Arterial	25A	1, 2	2, 10	32.469	<.001	Yes	Yes
	279A	2, 6	2, 10	15.811	<.001	Yes	Yes
	301A	1, 2	2, 10	21.530	<.001	Yes	Yes
	319A	1	1, 11	30.311	<.001	Yes	Yes
	42A	1	1, 9	9.965	.01-.025	Yes	Yes
	100X	1	1, 10	22.876	<.001	Yes	Yes
	256A	1	1, 11	18.899	.001-.005	Yes	Yes
	262A	1, 2	2, 10	5.309	.025-.05	Yes	Yes
Rural Major Collector	59A	1	1, 11	28.563	<.001	Yes	Yes
	200X	1	1, 8	13.940	.005-.01	Yes	Yes
	5420A	6	1, 11	23.447	<.001	Yes	Yes
	7047A	4	1, 11	11.988	.001-.005	Yes	Yes

(*) df_R = degrees of freedom for Regression.

df_F = degrees of freedom for Error.

Table 5.24

Partial F-tests for Disaggregate Analysis

Highway Category	ATR Station	Variable Subscripts for: Full Model	Variable Subscripts for: Reduced Model	df_R, df_F (*)	F^*	α	Is H_a true for α = .05?	Is H_a true for α = .10?
Rural Interstate	3070A	2, 8	2	11, 10	104.921	<.001	Yes	Yes
			8	11, 10	30.685	<.001	Yes	Yes
Rural Principal Arterial	134A	2, 7	2	11, 10	51.220	<.001	Yes	Yes
			7	11, 10	30.892	<.001	Yes	Yes
	173A	1, 2	1	11, 10	15.955	.001-.005	Yes	Yes
			2	11, 10	221.394	<.001	Yes	Yes
Rural Minor Arterial	25A	1, 2	1	11, 10	12.528	.005-.01	Yes	Yes
			2	11, 10	55.149	<.001	Yes	Yes
	279A	2, 6	2	11, 10	12.740	.005-.01	Yes	Yes
			6	11, 10	26.579	<.001	Yes	Yes
	301A	1, 2	1	11, 10	10.239	.005-.01	Yes	Yes
			2	11, 10	35.547	<.001	Yes	Yes
	262A	1, 2	1	11, 10	10.404	.005-.01	Yes	Yes
			2	11, 10	8.142	.01-.025	Yes	Yes

(*) df_R = degrees of freedom for SSE of Reduced Model.

df_F = degrees of freedom for SSE of Full Model.

be rejected) at 5 percent level of significance.

5.4.4 Model Development and Performance

The final regression equations are presented in Table 5.25, along with R^2 values, overall F values, t-statistics, and elasticities. The equations for the stations under rural interstates, rural principal arterials, rural minor arterials and rural major collectors explain 70.6 - 92.5, 60.7 - 95.5, 51.5 - 86.7 and 52.1 - 72.2 percent variation in AADT, respectively, by the use of the associated X-variable(s). Not all of the goals of Table 5.7 have been met in all of the equations in Table 5.25. However, the equations that resulted from the goals specified in Table 5.7 are the best possible, considering all the limitations. Using the elasticities obtained from the regression analysis, a forecasting model was developed for each station by substituting those elasticities into equation 3.1 (Chapter 3). These models are presented in Table 5.26. Each of the models is simple, with not more than two variables in any case. The use of these models is also straightforward. The data needed to predict rural traffic volumes with these models are readily available at the county, state and national levels. The models can be implemented with a hand-held calculator.

Table 5.25

Final Regression Equations from Disaggregate Analysis (*)

Rural Interstate			
Station 172A:	AADT = -74248.66 + 0.0171 State Population		
	R^2 = 0.706	t = 5.146	
	F = 26.478	e = 5.24231	
Station 3070A:	AADT = -105260.90 - 166.311 US Gas Price + 0.0241 State Population		
	R^2 = 0.925	t = -5.539	t = 10.243
	F = 61.951	e = -0.44503	e = 7.74428
Station 5474A:	AADT = -34973.79 + 1.1889 County Population		
	R^2 = 0.880	t = 8.962	
	F = 80.320	e = 6.18172	
Rural Principal Arterial			
Station 68A:	AADT = 924.99 + 0.0016 State Vehicle Registrations		
	R^2 = 0.955	t = 15.301	
	F = 234.112	e = 0.86979	
Station 134A:	AADT = 7120.83 - 116.511 US Gas Price + 0.0026 State Vehicle Registrations		
	R^2 = 0.837	t = -5.558	t = 7.157
	F = 25.611	e = -0.43949	e = 0.83878
Station 173A:	AADT = -2870.28 + 0.4903 County Vehicle Registrations - 52.2181 US Gas Price		
	R^2 = 0.978	t = 14.879	t = -3.994
	F = 222.058	e = 1.47643	e = -0.21371
Station 234B:	AADT = 2990.64 + 0.1495 County Vehicle Registrations		
	R^2 = 0.607	t = 4.125	
	F = 17.013	e = 0.60300	
Rural Minor Arterial			
Station 25A:	AADT = 1492.57 + 0.1286 County Vehicle Registrations - 24.9935 US Gas Price		
	R^2 = 0.867	t = 7.426	t = -3.540
	F = 32.489	e = 0.90147	e = -0.29365
Station 279A:	AADT = 4892.50 - 28.5409 US Gas Price + 0.0186 County Employment		
	R^2 = 0.760	t = -5.156	t = 3.569
	F = 15.811	e = -0.26635	e = 0.24526
Station 301A:	AADT = 2236.63 + 0.1423 County Vehicle Registrations - 22.2131 US Gas Price		
	R^2 = 0.812	t = 5.962	t = -3.200
	F = 21.530	e = 0.66731	e = -0.26576

Table 5.25 (continued)

Rural Minor Arterial			
Station 319A:	AADT = 752.47 + 0.0471 County Vehicle Registrations		
	R^2 = 0.734	t = 5.506	
	F = 30.311	e = 0.61456	
Station 42A:	AADT = 2147.18 + 0.0279 County Vehicle Regigtrations		
	R^2 = 0.525	t = 3.157	
	F = 9.965	e = 0.49887	
Station 100X:	AADT = 3045.80 + 0.0168 County Vehicle Registrations		
	R^2 = 0.696	t = 4.783	
	F = 22.876	e = 0.64875	
Station 256A:	AADT = 1861.50 + 0.0819 County Vehicle Registrations		
	R^2 = 0.632	t = 4.347	
	F = 18.899	e = 0.33059	
Station 262A:	AADT = 2371.90 + 0.0356 County Vehicle Registrations − 12.9828 US Gas Price		
	R^2 = 0.515	t = 2.853	t = −3.226
	F = 5.309	e = 0.28236	e = −0.23256
Rural Major Collector			
Station 59A:	AADT = 2771.53 + 0.0481 County Vehicle Registrations		
	R^2 = 0.722	t = 5.344	
	F = 28.563	e = 0.36063	
Station 200X:	AADT = 6557.79 + 0.0503 County Vehicle Registrations		
	R^2 = 0.635	t = 3.734	
	F = 13.940	e = 0.26407	
Station 5420A:	AADT = 784.34 + 0.1163 County Employment		
	R^2 = 0.681	t = 4.842	
	F = 23.447	e = .59744	
Station 7047A:	AADT = 1122.06 − 0.0433 County Population		
	R^2 = 0.521	t = −3.462	
	F = 11.988	e = −3.48274	

(*) For unit and symbol of each variable, see Table 4.1 of Chapter 4.

Table 5.26

Disaggregate Traffic Forecasting Models (*)

	Rural Interstate
Station 172A:	$AADT_f = AADT_p$ [1 + 5.24231 (Δ State Population)]
Station 3070A:	$AADT_f = AADT_p$ [1 − 0.44503 (Δ US gas Price) + 7.74426 (Δ State Population)]
Station 5474A:	$AADT_f = AADT_p$ [1 + 6.18172 (Δ County Population)]
	Rural Principal Arterial
Station 66A:	$AADT_f = AADT_p$ [1 + 0.86979 (Δ State Vehicle Registrations)]
Station 134A:	$AADT_f = AADT_p$ [1 − 0.43949 (Δ US gas Price) + 0.83878 (Δ State - Vehicle Registrations)]
Station 173A:	$AADT_f = AADT_p$ [1 + 1.47643 (Δ County Vehicle Registrations) − 0.21371 (Δ US - Gas Price)]
Station 254B:	$AADT_f = AADT_p$ [1 + 0.60300 (Δ County Vehicle Registrations)]
	Rural Minor Arterial
Station 25A:	$AADT_f = AADT_p$ [1 + 0.90147 (Δ County Vehicle Registrations) − 0.29065 (Δ US - Gas Price)]
Station 279A:	$AADT_f = AADT_p$ [1 − 0.26635 (Δ US gas Price) + 0.24526 (Δ County Employment)]
Station 301A:	$AADT_f = AADT_p$ [1 + 0.66731 (Δ County Vehicle Registrations) − 0.26576 (Δ US - Gas Price)]
Station 319A:	$AADT_f = AADT_p$ [1 + 0.61456 (Δ County Vehicle Registrations)]

Table 5.26 (continued)

Rural Minor Arterial	
<u>Station 42A:</u>	$AADT_f = AADT_p$ [1 + 0.49887 (Δ County Vehicle Registrations)]
<u>Station 100X:</u>	$AADT_f = AADT_p$ [1 + 0.64875 (Δ County Vehicle Registrations)]
<u>Station 256A:</u>	$AADT_f = AADT_p$ [1 + 0.33059 (Δ County Vehicle Registrations)]
<u>Station 262A:</u>	$AADT_f = AADT_p$ [1 + 0.28238 (Δ County Vehicle Registrations) - 0.23256 (Δ US - Gas Price)]
Rural Major Collector	
<u>Station 59A:</u>	$AADT_f = AADT_p$ [1 + 0.36063 (Δ County Vehicle Registrations)]
<u>Station 200X:</u>	$AADT_f = AADT_p$ [1 + 0.28407 (Δ County Vehicle Registrations)]
<u>Station 5420A:</u>	$AADT_f = AADT_p$ [1 + 0.59744 (Δ County Employment)]
<u>Station 7047A:</u>	$AADT_f = AADT_p$ [1 - 3.48274 (Δ County Population)]

(*) (i) For unit and symbol of each variable, see Table 4.1 of Chapter 4.

(ii) Δ represents change in predictor variable with respect to its present value in fraction.

For example, $\Delta X = \dfrac{X_f - X_p}{X_p}$, where X_p & X_f represents present and future value of X.

The ability of the models shown in Table 5.26 to predict 1983 and 1984 traffic volumes was tested using 1980 as the "present year". The 1983 and 1984 data were not used in the development of the model, but now can be used to allow a comparison of the accuracy of the disaggregate model with extrapolation. The results of this comparison are shown in Table 5.27. Table 5.27 also shows AADT for years 1983 and 1984 obtained from simple extrapolation. Figures F1 to F4, selected randomly from the 19 figures in Appendix F, illustrate how this extrapolation is carried out. In these figures, an average line is drawn for each plot through the data points and is then extrapolated to 1984. This simple extrapolation is a very crude method. But, Table 5.16 shows that simple extrapolation often gives better results over the short-range with aggregate models. This simple extrapolation will not likely provide good results over longer ranges (more than 10 years). While the proposed model is expected to provide better results because it is based on the functional relationship between the response variable (AADT in this case) and predictor variable(s).

The disaggregate model's forecasts come closer to the actual "future values" than the extrapolations in a majority of the cases. The prediction errors for either method are not more than 15 percent. In general, both the simple extrapolation and the disaggregate models provide

Table 5.27

Performance of Disaggregate Traffic Forecasting Models

Highway Category	Station	Year	Actual AADT	Predicted AADT	Prediction Error in percent (*)	Extrapolated AADT	Extrapolation Error in percent (*)
Rural	172A	1983	18454	18454	.49	20750	12.44
		1984	19091	18885	-1.08	21000	10.00
Interstate	3070A	1983	18219	19171	5.22	19800	8.68
	5474A	1983	7047	7161	1.62	8000	13.52
		1984	7541	7281	-3.45	8050	6.75
Rural	68A	1983	7969	7642	-4.10	8100	1.64
		1984	8105	7816	-3.57	8200	1.17
Principal	134A	1983	12366	12765	3.23	13200	6.74
	173A	1983	12751	12087	-5.21	12900	1.17
Arterial	254B	1983	9031	8086	-10.46	8800	-2.56
		1984	9661	8244	-14.87	8950	-7.36
Rural	25A	1983	4245	4136	-2.57	4320	1.77
	279A	1983	4762	5144	8.02	5020	5.42
Minor	301A	1983	3793	4238	11.73	4140	9.15
Arterial	319A	1983	2211	2201	-.45	2420	9.45
		1984	2279	2236	-1.89	2460	7.94

Table 5.27 (continued)

Highway Category	Station	Year	Actual AADT	Predicted AADT	Prediction Error in percent (*)	Extrapolated AADT	Extrapolation Error in percent (*)
Rural	42A	1983	4515	4353	-3.59	4675	3.54
		1984	4607	4411	-4.25	4710	2.24
	100X	1983	9103	8608	-5.44	9420	3.48
Minor		1984	9560	8643	-9.59	9540	-.21
	256A	1983	2861	2839	-.04	3019	5.21
		1984	2940	2671	-2.45	3020	2.62
Arterial	262A	1983	2488	2617	5.18	2620	5.31
Rural	59A	1983	4551	4701	3.30	4780	5.03
		1984	4769	4718	-1.07	4800	1.73
	200X	1983	9297	9471	1.87	9700	4.33
Major		1984	9950	9566	-3.84	9760	-1.91
	5420A	1983	1979	2360	6.97	2340	18.24
	7047A	1983	281	262	-6.76	302	7.47
Collector		1984	273	262	-4.03	310	13.55

(*) '+' sign indicates overprediction and
'-' sign indicates underprediction.

comparable forecast errors in this short range of time. But, it is expected that the disaggregate model will provide increasingly better traffic forecasts than extrapolation as the planning horizon increases, while the projections from extrapolation will lose accuracy. While this short range comparison between disaggregate models and extrapolation is inconclusive, there is an indirect indication that disaggregate models perform better over this time span than aggregate models. This compatible with the comparative statistical measures obtained during the development and refinment of both model times. In general, with a lower number of variables, the disaggregate models yielded lower prediction error than the aggregate models (See Tables 5.16 and 5.27).

CHAPTER 6

SUMMARY AND CONCLUSIONS

Both aggregate and disaggregate traffic forecasting models models for rural state highways in Indiana were developed using traffic data from Automatic Traffic Record (ATR) stations and economic and demographic variables for the county, state and national levels. The models and the described procedure are intended to provide highway planners with a tool for simple, fast and inexpensive estimation of traffic projections. Some problems and limitations of the models and suggestions to overcome the problems have been discussed. This chapter presents the steps to implement the models and makes recommendations for further studies.

6.1 Guidelines for Applicability of Models

Preliminary statistical analysis (Chapter 5) favored the disaggregate model applied to each station separately

over the aggregate model applied to each highway class. The disaggregate models are location-specific, but the aggregate models are general in nature for a particular highway category. However, the use of disaggregate models is not limited only to the locations for which they are developed. If a project site, for which a forecast of future traffic is needed, can be shown to be "similar" to a station for which a disaggregate model has been built, then the disaggregate model of the station could be employed. The following points are provided as a guide deciding whether a section of highway is "similar" to a station for which a disaggregate model has been developed:

1. The statistical test for equality of two population means could be carried out for the response (Y) and predictor variables (X´s) at the county level to see if the mean of these variables are the same for the two points or section of highway under consideration. The hypothesis and the decision rule for this test are explained in Appendix G.

2. The stage of commercial and industrial land development, measured as a percentage of commercial and industrial land to the total land, of the two counties should be approximately similar.

3. The highway type, its geographical location with respect to traffic generators (for example, schools,

hospitals, restaurants, shopping centers, etc.), and road network characteristics of the two points should be similar.

The aggregate model is general in nature for a particular highway category. The aggregate model for a category of highway is designed to be applicable for any section under that category of highway, although it is usually not as reliable as the disaggregate model. If a project site can not be shown to be "similar" to a station, then the aggregate model should be applied to that site.

6.2 Summary of Aggregate Models

Elasticity-based aggregate traffic forecasting models were presented in Table 5.15. Each of these models is simple and does not contain more than three predictor variables (X's) in it. The models have good R^2 (65.8 percent to 83.7 percent) values. These models are statistically sound and simple, with only one predictor variable in three cases and two predictor variables in the other case. The results of the performance of these models were presented in Table 5.16 for the stations not used in model development. The forecasted errors are reasonably small in most of the cases and speak well for the reliability of the models. The choice of predictor variables for the models was based on the combination of statistical analysis and subjective judgment. The

predictor variables used in the models were found significant at the 5 percent level of significance. (See Tables 5.12 and 5.13.) The resulting models were found to be satisfactory within the limitation of data available.

6.3 Summary of Disaggregate Models

Elasticity-based disaggregate models were presented in Table 5.26. Each of the models is simple and does not contain more than two predictor variables. The models have good R^2 values -- 51.5 percent to 97.8 percent. The results of the performance of the models (presented in Table 5.27) showed that the prediction/forecasting errors in 88 percent of the cases were found to be equal to or less than 10 percent. The larger prediction errors (more than 10 percent) in the rest of the cases are due to insufficient data. The choice of predictor variables for the models was based on a combination of statistical analysis and subjective judgment, as described in Sections 5.3.2.1 to 5.3.3.1. The predictor variables used in the models were found significant at as low as the 5 percent level of significance (see Tables 5.23 and 5.24). The disaggregate models were found to be satisfactory within the limitation of data and better than the aggregate models with respect to performance and graphic residual analysis.

6.4 Problems, Limitations and Suggestions

A few problems may appear as soon as users begin to use the models to predict rural traffic. The most serious problem in the application of this procedure is one that is common to all forecasting processes: the accuracy of the model is determined to a large extent by the accuracy of the input, especially the future values of the predictor variables (X's). In this study, the following predictor variables were used in disaggregate models: (1) population, (2) households, (3) vehicle registrations, (4) employment and (5) gas price. On the other hand, aggregate models were developed using only (1) population and (2) households. The Indiana University Business School [26] projects the population and number of households for every fifth year into the future, but there is very little information available for the other variables required by the disaggregate models. The question then is how to estimate future values for vehicle registrations, employment and gas price.

Several options could be suggested to obtain future estimates of vehicle registrations. The first, and most appropriate, is to check the Bureau of Motor Vehicles to see if they have forecasts appropriate for our model. If that fails, then the following methods [38] could be employed to forecast future vehicle registrations:

1. Calculate the average annual growth rate from the historical data (say 1970 to 1982 data, which were used in data tables), and assume an increasing, decreasing, or constant rate for the future. This method does not consider reaching a saturation level of vehicle ownership, but it may be reflected by altering the projected growth rate.

2. The saturation phenomenon that could be employed to estimate future vehicle registrations is the only difference in this method from the first method, described above. Examine the trend of vehicles per person in the previous years and then carry that trend out to the future until the value reaches a pre-defined saturation level. For example, the trend of vehicle per person for the State of Indiana is shown in Figure 6.1. From this figure, 0.85 could be taken as the saturation level of vehicles per person for the State of Indiana. Then, by multiplying the projected number of vehicles per person in a future year by the population forecast, an estimate of that year's vehicle registrations can be obtained.

The ways to obtain the future values for employment are similar to those of vehicle registrations. First, and most appropriate, is to contact the Employment Security Division [24] for employment forecasts. If that fails,

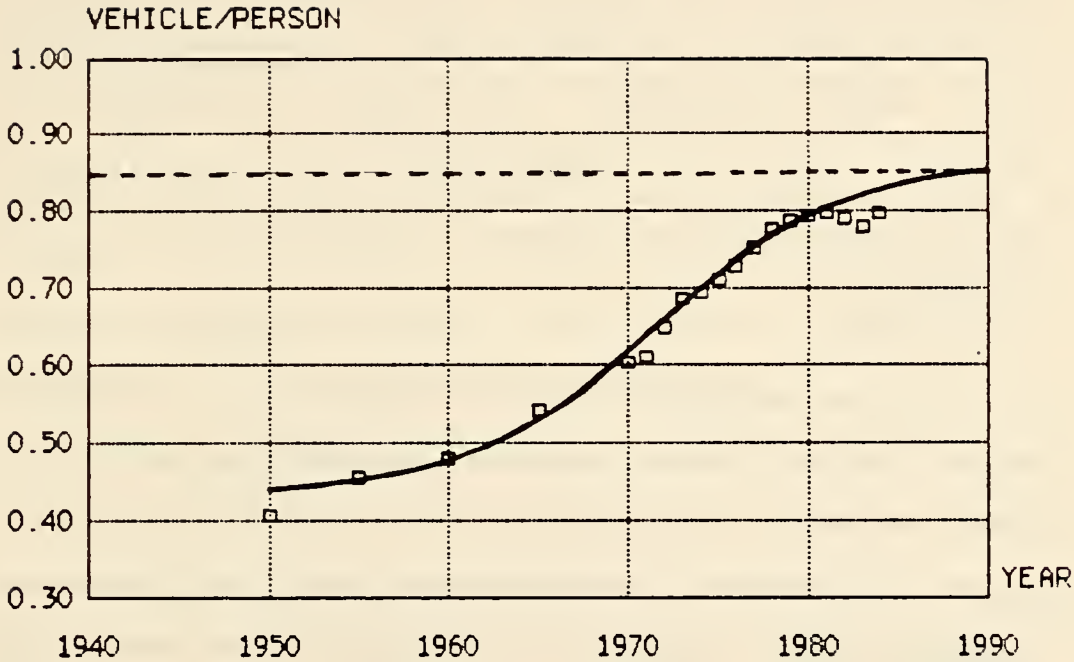


Figure 6.1: Trend of Vehicle/Person, State of Indiana

then the easiest way is to calculate the average annual growth rate from historical data (1970 to 1982 data were used in data tables), and assume an increasing, decreasing, or constant rate for the future. Also, the local employment office may be able to provide some information about the future levels of employment. A second method is to calculate employment per person in the previous years. In this method, the saturation level, if any, could be employed. The trend is carried out to future. Then, by multiplying this trend by estimated population in future years, employment data could be developed.

To get future values of US gas price, a first step will be to check whether the Independent Petroleum Association of America and/or US Department of Energy have useable forecasts. If no outside fuel price forecasts are available, the user still has recourses. He can devise a series of simple fuel prices projections (by extrapolation, etc.) to produce a range of values that can represent high, medium, and low fuel price scenarios. The results of these values used in traffic forecasting models can then be compared with the results of models that do not require fuel price as an input variable, if such models exist. At a minimum, the traffic forecasts based on the range of fuel price values could be compared against a range of traffic volume extrapolations, in

search of some degree of consensus.

Applicability of the models in various areas may also cause problems. How can a user decide whether the project area, for which future forecast of traffic is needed, is "rural" enough for the model(s)? It is difficult to provide guidelines to assist in this issue, but suggestions [49] exist. Judgment is required in making this determination. In very approximate terms, highways with more than 10 uncontrolled access points per mile (on one side) would be considered to be "suburban". Also, any highway on which left or right turns cause appreciable delay to through vehicles would also be classified as "suburban". Multilane suburban highways and rural roads differ from suburban arterials in the following features: (1) their roadside development is not as intense, (2) the density of traffic access points is not as high, and (3) signalized intersections are more than 2 miles apart. In fact, highways with signal spacing of 2 miles or more could be treated as "rural" highways. Increased use of the developed models will lessen this problem.

The model formulation in Chapter 3 assumes that elasticities are constant over time. Historically, travel has been growing at a fairly constant rate for many years. Although fuel shortages interrupted this increasing rate for a while, it has resumed. Therefore, any assumption of constant elasticities would not introduce substantial

errors. On the other hand, variable elasticities are not very common in traffic forecasting, which involve more sophisticated and expensive analysis [28]. The sophisticated and expensive analysis is against the principles that the models should be easy to understand and less costly. But, when new census data become available, the elasticities could be recalculated and the appropriateness of earlier values could be checked. If the elasticities seem to change significantly (for example, more than 10 percent), then the new set of elasticities should be used in the model.

Users are expected to weigh the results of forecasting models in terms of the local situation, and adjust them according to their professional judgment of the specific area.

6.5 Stepwise Plan for Implementation

The steps that are recommended for the implementation of the aggregate and disaggregate models to predict the future traffic for rural roads of Indiana are listed below.

1. Determine the exact location (i.e., county) of the roadway for which forecast is needed.

2. Select the traffic model(s) that will be used to predict traffic.

a. Determine the functional class of roadway.

This will determine which aggregate model is applicable to the project site. To determine the functional class, the functional classification system map, prepared by division of planning of Indiana State Highway Commission, will be the best guide. Moreover, the definitions provided in Section 2.8 of this report would be helpful to find the appropriate highway class. The project site will be classified in one of the four categories of highways provided in Table 4.2. If the classification is not clear-cut, then personal judgment should be used, and documentation provided.

b. Examine the project site with respect to ATR stations.

Check if the project site is one of the Automatic Traffic Record (ATR) stations, used in the development of models. If it is one of the stations used in the model development, then the disaggregate model for that station will be applicable. Otherwise, the procedures described in Section 6.1 could be used to classify determine if the project site is "similar" to

one of the stations used in the model development. In case the project site is not found to be "similar" to a ATR station, the aggregate model for a highway category must be applied to the project site. Identification of the highway category of the project site is the only criterion used to select the appropriate aggregate model.

3. Collect the base year AADT.

The base year AADT of the project site can be one value for a small project (e.g., intersections), or a series of estimates for roadway sections for a larger project (e.g., lane widening). One possible source of data would be the Highway Department's Traffic Volume Book. If the Traffic Volume Book fails to provide such information, then it could be determined from short-term counts at the project site, using the procedure described in Section 2.6.

4. Collect the base and future year data for the predictor variables.

The description of variables in Section 4.2 is a guide to the sources of the required predictor variables. Section 6.4 will also be helpful, particularly with reference to future year data for the required predictor variables.

5. Estimate the future year AADT.

   a. Calculate the future year AADT by using the appropriate aggregate model (Table 5.15), as determined in step 2(a), with the values found in steps 3 and 4. Denote this AADT estimate as $AADT_a$.

   b. Calculate the future year AADT by using appropriate disaggregate model (Table 5.26), as determined in step 2(b), if possible, with the values found in steps 3 and 4. Denote this AADT estimate as $AADT_d$.

   c. Find the weighted average of the two AADT estimates found in steps 5(a) and 5(b). The users may give more weight to the AADT found in step 5(b), because it was found that the disaggregate model performs better than the aggregate model. The weighted average of AADT is calculated by using equation 6.1.

$$AADT_w = w * AADT_a + (1 - w) * AADT_d \qquad (6.1)$$

where,

w = Weight given to AADT estimate done by aggregate model, $0 < w < 1.00$,

$AADT_a$ = AADT estimate by aggregate model,

$AADT_d$ = AADT estimate by disaggregate model,

$AADT_w$ = weighted AADT estimate.

In general, given the better performance of disaggregate models with respect to aggregate models, the value of w is recommended to be less than 0.50 (users are suggested to use a value of w between 0.35 to 0.45). If an AADT estimate using disaggregate model (step 5b) is not possible, then the value of w must be 1.

6. Adjust the estimated future year AADT.

If historical AADT counts are available for the project site, plot AADT against time and extend the trend to the future year. Check whether the projected AADT differs significantly (say, more than 25 percent) from the AADT estimate found at step 5. In case of a significant difference, an average of the estimate at step 5 and the extension of plot of AADT against time at the desired year may be taken as the "future year AADT". Otherwise, the estimate

result after step 5 will be the "future year AADT".

6.6 Recommendations for Future Study

The methodology presented in this report was based on a small number of continuous count stations. The aggregate and disaggregate traffic forecasting models for rural roads of Indiana were developed using this methodology. Continuous count stations are the only locations where "true" historic AADT counts are available. Further traffic forecasting studies will be helped by the installation of more continuous count stations at locations representing a variety of highway categories and traffic characteristics. It is expected that, with an increased number of continuous count stations, the present methodology will provide better statistical results and model performance. Moreover, with an increased number of count stations, it may become possible to divide the whole rural state network into regions or otherwise separate different historical growth rates. Statistical methods could be employed to identify the different sectors or groupings. In the early stage of this study, this approach was attempted, but dropped due to the limited number of ATR stations. The development of a model for each sector would be similar to aggregate and disaggregate models developed in this report.

Time series analysis could be used to forecast future traffic. According to Armstrong [4,5], the time series approach could be combined with the present approaches to obtain reliable traffic forecast. Time series analysis treats traffic volume as a function of time and uses land use development as the starting point to formulate the traffic growth -- as time passes, more land is developed and traffic increases proportionally. Time series analysis is also a way to introduce time lags, especially with respect to economic predictor variables, to see if better ADT forecasting models are possible.

The variables used in statistical analysis are more or less subject to error. The prediction in this case could be further modified by introducing an error term in the regression formulation. Prediction considering error-in-variables is not well practiced. Ganse et al. [50] made predictions of earthquake magnitudes by employing the consideration of error-in-variables.

One of the major problems encountered in aggregate analysis was "mix-normal" data. The AADT data for each station has a normal distribution. But, when the stations were combined in aggregate analysis as a highway category, the AADT data failed to produce a normal distribution, primarily due to the limited number of count stations. The treatment of this mixture distribution is also a new area is statistical science. Kotz et al. [51] provided

some useful theoretical discussion of this mixture distribution. Exhaustive investigations failed to find any treatment regarding "mix-normal" that could be directly employed in prediction. In the absence of suitable computer program(s), available program(s) could be modified using the theory under mixture treatment. If this "mix-normal" problem in the aggregate analysis is solved, then the aggregate models will provide better results than the results found in the present study. In that case, it will also reduce the necessity to increase the number of count stations.

6.7 Conclusions

The principal objective of this report was to develop simple, fast and inexpensive traffic forecasting models for rural state highways in Indiana. The study first identified suitable methodologies and then applied statistical analyses to find suitable variables to employ in the models. The analyses done to develop the elasticity-based aggregate and disaggregate models are as reliable as possible within the limitations of the data. The developed models could be updated as new data become available. The developed models provide better statistical results (for example, R^2) than those found in a previous, similar study [38]. Moreover, variable selection criteria used in this study are not based solely

on stepwise regression. The variables used in the models were found statistically significant and it was found that no other variables will provide additional significant predictive power in the models.

The step-by-step instructions in Section 6.4 are provided to give a structured approach in implementing the models. The developed models are expected to provide a means to highway planners for simple, fast, and inexpensive estimation of future traffic.

In almost every state, the task of traffic forecasting for the rural areas is heavily dependent on the AADT counts at continuous count stations. Any state with adequate historical traffic data at continuous count stations could employ this model building approach to determine future year AADT at rural locations.

The prediction of rural traffic volume has been relatively neglected despite its many potential uses. The most obvious and direct use of rural traffic forecasting model is for the estimation of the benefits from alternate highway system improvement projects. A second application would be as an aid to the appropriate design of a project (for example, number of lanes or type of traffic control). The identification of potential problem segments in the state highway system could be accomplished by using the models to identify rapid traffic growth areas.

Undoubtedly, more work must be done in this area to improve the accuracy and reliability of a traffic projection model. It is important to note that the developed models in this report are not purported to be perfect forecasting tools, if such a model could ever exist. Users are expected to weigh the results in terms of the local situation, and make adjustments in accordance with their professional judgment. Finally, it is expected that combining different methods will provide more reliable traffic forecasts to highway planners.

LIST OF REFERENCES

LIST OF REFERENCES

1. AASHTO, A Policy on Geometric Design of Highways and Streets, 1984.

2. Albright, D., "Abstract on - Trend-Line: Forecasting Heavy Commercial Traffic as a Percent of Average Daily Traffic", New Mexico State Highway Department, 1985.

3. Anderson, V. L. and McLean, R. A., Design of Experiments: A Realistic Approach, Statistics: text books and monographs, volume 5, Marcel Dekker, Inc., New York, U.S.A., 1974.

4. Armstrong, J. Scott, "The Ombudsman: Research on Forecasting: A Quarter-Century Review (1960-1984)", Interfaces, 16:1, Jan.-Feb. 1986, pp.89-109.

5. Armstrong, J. Scott, "Forecasting by extrapolation: Conclusion from 25 years Research", Interfaces, 14:6, Nov.-Dec. 1984, pp.52-56.

6. Benjamin, J., "A Time Series Forecast of Average Daily Traffic Volume", Economics Department and Transportation Institute, North Carolina A & T State University , Greensboro, North Carolina, 1984.

7. Boyce, R. F., "Cohort Population Forecasting Program (7 parts) written in Basic for Apple IIe using DOS 3.3", Planning Bureau, New Mexico State Highway Department, 1985.

8. Bureau of Labor Statistics, "Handbook of Labor Statistics", Bulletin 2217, U.S. Department of Labor, June 1984.

9. Bureau of the Census, "County and City Data Book (1956, 1962, 1967, 1972, 1977, 1983): A Statistical Abstract Supplement", U.S. Department of Commerce.

10. Bureau of the Census, "County Business Patterns - Indiana, 1970 - 1983", U.S. Department of Commerce.

11. Bureau of the Census, "Characteristics of the Population, Number of Inhabitants, Indiana, 1980 Census of Population", U.S. Department of Commerce.

12. Chen, S. Hong, "Improved Methods for State-wide Vehicle Counting", JHRP-81-6, Purdue University and Indiana State Highway Commission, 1981.

13. Council of Economic Advisers, Government Printing Office, Washington, D.C., "Economic Indicators" - A monthly Publications, Prepared for the Joint Economic Committee, May 79, April 80, June 84, Dec. 85.

14. Covault, O. D., "Indiana's Highway needs Study", HRB Bulletin 194, Jan. 1958, pp. 14-27, JHRP, Purdue University.

15. Drapper, N. R. and Smith, H., Applied Regression Analysis, John Wiley and Sons, New York, 1981.

16. Drusch, R. L., "Estimating Annual Average Daily Traffic from Short Term Traffic Counts", HRB Record 118, 1966, pp. 85-95.

17. Economic Statistics Bureau of Washington, D.C., The Handbook of Basic economic Statistics - Labor, Industry, Commerce and Agriculture; April 1984, Vol. XXXXIX , No. 4, pp. 99-101.

18. Federal Highway Administration, "Guide for Traffic Volume Counting Manual", Bureau of Public Roads, U.S. Department of Transportation, March 1970.

19. Freund, R. J. and Minton, P. D., Regression Methods - A Tool for Data Analysis, Marcel Dekker, Inc., New York, 1979.

20. Hartgen, D. T., "What will happen to travel in the next 20 years?", Transportation Research Record 807, TRB, Dec. 1981, pp. 21-27.

21. Hazen, P. I., "Summary of Recent National and State Highway Travel Forecast Studies", U.S. Department of Transportation, Program Management division, Federal Highway Administration, January 1982.

22. Indiana Bureau of Motor Vehicles, "Motor Vehicle Registration in Indiana By County (1955 - 1984)".

23. Indiana Department of Highways (IDOH), "Traffic Statistics - 1982", Division of Planning, IDOH.

24. Indiana Employment Security Division (IESD), "County Employment Patterns, 1970 - 1983", Research and Statistics Section, IESD.

25. Indiana State Highway Commission (ISHC), "Traffic Flow Map (1952, 1958, 1962, 1966, 1969, 1972, 1975) -- 24 hr AADT", Division of Planning, ISHC.

26. Indiana University, "Preliminary Estimates of the Intercensal Population of Counties in Indiana", School of Business, Indiana Business Research Center, Indiana University.

27. ITE Committee 6A-1, "Estimating ADT on Low Volume Roads(less than 2000 vpd), Traffic Engineering, vol. 44, No. 13, Oct. 1974, pp. 19-22.

28. Kanafani, A., Transportation Demand Analysis, McGraw-Hill Book Company, 1983.

29. Magridge, M. J. H., "Traffic Forecasting - Part I", Transport and Road Research Laboratory (TRRL), V45, n1811, Aug. 1977, pp. 5-10.

30. Magridge, M. J. H., "Traffic Forecasting - Part II", TRRL, V45, n1812, Oct. 1977, pp. 5-9.

31. Mckay, J. G., "Population and Highway Traffic", Highway Research Board Proceedings, Vol 5, pp. 192-197.

32. Memmott, J. L. and Buffington, J. L., "Predicting Traffic Volume Growth rates resulting from changes in Highway Capacity and Land Development", Texas Transportation Institute, Texas A&M University, 1981.

33. Memmott, J. L., "Factors that Affect Growth Rates and Projections of Traffic Volumes for Use in Highway Economic Models", Transportation Research Record 912, 1983.

34. Meyer, M. D. and Miller, E. J., Urban Transportation Planning: A Decision - Oriented Approach, McGraw-Hill Book Company, 1984.

35. Minnesota Department of Transportation (MnDOT), "Procedures Manual for Forecasting Traffic on the Rural Trunk Highway System", Traffic Forecasting Unit, Program Management Division, MnDOT, April 1985.

36. Morf, F. T. and Houska, V. F., "Traffic Growth Pattern on Rural Areas", Highway Research Board Bulletin 194, January 1958, pp. 33-41.

37. Neter, J. and Wasserman, W., Applied Linear Statistical Models, Richard D. Irwin, Inc., Illinois, 1974.

38. Neveu, A. J., "Quick Response Procedures to Forecast Rural Traffic", Transportation Research Record 944, June 1982.

39. Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K. and Bent, D.H., SPSS - Statistical Package for the Social Sciences, Second Edition, McGrew-Hill Book Company.

40. Petroff, B. B. and Blensly, R. C., "Improving Traffic Count Procedures by Applications of Statistical Methods", HRB Proceedings, Vol. 33, 1954, pp. 362-375.

41. Petroff, B. B., "Some Criteria for Scheduling Mechanical Traffic Counts", HRB Proceedings, Vol. 26, 1946, pp. 389-396.

42. Purdue University Computer Center (PUCC), "DRRSQU (Program for C_p-statistic - All possible Regression): Statistical Program", Statistical Library, PUCC, Purdue University, West Lafayette, Indiana.

43. Ritchie, S. G., "A Statistical Approach to Statewide Traffic Counting", Transportation Research Board Annual Meeting, Washington, Jan. 1986.

44. Salovara, V., Makela, O. and Koljonen, K., "Forecast of Traffic and Number of Motor Vehicles (1980-2000)", TRRL, IRRD 248155.

45. Standard and Poor´s - Industry Surveys, "Oil Basic Analysis", Section 2, Nov. 1984, pp. 74.

46. Taliadoros, C. A., "Determining Growth Patterns at the Continuous Count Stations", Thesis presented to Department of Civil Engineering, Purdue University for MSCE, Dec. 1983.

47. University of California, Los Angels (UCLA), Biomedical Computer Programs (BMDP): P-series, Health Science Computing facility, UCLA, 1979.

48. Winer, B. J., Statistical Principles in Experimental Design, Second edition, McGraw-Hill Book Company, 1971, pp. 876.

49. Transportation Research Board (TRB), "Highway Capacity Manual", Special Report 209, TRB, National Research Council, Washington, D.C., 1985.

50. Ganse, R. A., Amemiya, Y. and Fuller, W. A., "Prediction When Both Variables are Subject to Error, With Application to Earthquake Magnitudes", Journal of the American Statistical Association, Dec. 1983, Volume 78, Number 384.

51. Kotz, S. and Johnson, N. L. (Editors-in-Chief), Encyclopedia of Statistical Sciences, Volume 5, John Wiley & Sons, 1985.

APPENDICES

Appendix A

Data Tables
for
Aggregate Analysis

Table A1

Data Table for Rural Interstate

Station County	a	b	c	d	e	f	g	h	i	j	k	l	m	n
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	13271.	22447.	37.97	1970.	33187.	10583.	7752.	3136244.	5195392.	1609443.	1580100.	116.3	1085.6	3619.
	14239.	22822.	36.88	1971.	33000.	10631.	7601.	3210180.	5250000.	1645684.	1531495.	121.3	1122.4	3714.
	16126.	24326.	36.13	1972.	33400.	10870.	8513.	3439698.	5296000.	1680066.	1580411.	125.3	1185.9	3860.
	17942.	25743.	37.60	1973.	34000.	11181.	9104.	3652567.	5329000.	1711111.	1702849.	133.1	1254.3	4080.
	17047.	26376.	46.32	1974.	34300.	11398.	9105.	3716215.	5350000.	1739021.	1716132.	147.7	1246.3	4009.
	17746.	26679.	45.14	1975.	34600.	11620.	7873.	3772173.	5351000.	1761044.	1603117.	161.2	1231.6	4051.
172A	18455.	27657.	42.75	1976.	34700.	11779.	8539.	3909522.	5372000.	1790290.	1661631.	170.5	1298.2	4158.
Jackson	18823.	28773.	42.09	1977.	35000.	12010.	10219.	4071356.	5405000.	1824333.	1727408.	181.5	1369.7	4280.
	19879.	30370.	40.46	1978.	35400.	12281.	11900.	4238839.	5446000.	1861995.	1839974.	195.4	1438.6	4441.
	19679.	31280.	48.47	1979.	36000.	12628.	12792.	4319035.	5475000.	1896489.	1929580.	217.4	1479.4	4512.
	18745.	32000.	59.45	1980.	36523.	12955.	12280.	4362862.	5490180.	1927050.	1846400.	246.8	1475.0	4487.
	17772.	32523.	56.40	1981.	37000.	13273.	12170.	4374142.	5489000.	1952615.	1788511.	272.4	1512.2	4561.
	17801.	32401.	51.63	1982.	37300.	13535.	11783.	4342071.	5482000.	1976781.	1719683.	289.1	1480.0	4555.
	12999.	23174.	37.97	1970.	35096.	10792.	3676.	3136244.	5195392.	1609443.	1580100.	116.3	1085.6	3619.
	14557.	24314.	36.88	1971.	35200.	10896.	3935.	3210180.	5250000.	1645684.	1531495.	121.3	1122.4	3714.
	16060.	26423.	36.13	1972.	36700.	11437.	4756.	3439698.	5296000.	1680066.	1580411.	125.3	1185.9	3860.
	16914.	28570.	37.60	1973.	38400.	12048.	5359.	3652567.	5329000.	1711111.	1702849.	133.1	1254.3	4080.
	15898.	29951.	46.32	1974.	39600.	12509.	5480.	3716215.	5350000.	1739021.	1716132.	147.7	1246.3	4009.
	16437.	31284.	45.14	1975.	40100.	12754.	5256.	3772173.	5351000.	1761044.	1603117.	161.2	1231.6	4051.
3070A	17433.	32485.	42.75	1976.	40700.	13034.	5588.	3909522.	5372000.	1790290.	1661631.	170.5	1298.2	4158.
Hancock	18329.	35745.	42.09	1977.	41400.	13370.	5889.	4071356.	5405000.	1824333.	1727408.	181.5	1369.7	4280.
	18818.	37448.	40.46	1978.	42100.	13672.	7981.	4238839.	5446000.	1861995.	1839974.	195.4	1438.6	4441.
	17517.	38695.	48.47	1979.	43200.	14128.	8397.	4319035.	5475000.	1896489.	1929580.	217.4	1479.4	4512.
	17211.	38017.	59.45	1980.	43939.	14472.	8432.	4362862.	5490180.	1927050.	1846400.	246.8	1475.0	4487.
	17095.	38234.	56.40	1981.	43900.	14562.	8038.	4374142.	5489000.	1952615.	1788511.	272.4	1512.2	4561.
.	17963.	38447.	51.63	1982.	43800.	14634.	7769.	4342071.	5482000.	1976781.	1719683.	289.1	1480.0	4555.

a - y	e - X_4	i - X_8	l - X_{11}
b - X_1	f - X_5	j - X_9	m - X_{12}
c - X_2	g - X_6	k - X_{10}	n - X_{13}
d - X_3	h - X_7		

Note: For the meaning and definition of each variable, see Table 4.1 and Chapter 4 in the text.

Table A2

Data Table for Rural Principal Arterial

Station County	a	b	c	d	e	f	g	h	i	j	k	l	m	n
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
	5703.	20234.	37.97	1970.	29430.	9025.	6201.	3136244.	5195392.	1609443.	1580100.	116.3	1085.6	3619.
	6127.	19533.	36.88	1971.	29300.	9062.	6474.	3210180.	5250000.	1645684.	1531495.	121.3	1122.4	3714.
	6491.	20989.	36.13	1972.	30100.	9389.	7286.	3439698.	5296000.	1680066.	1580411.	125.3	1185.9	3860.
	6867.	22161.	37.60	1973.	30400.	9565.	7532.	3652567.	5329000.	1711111.	1702849.	133.1	1254.3	4080.
	6683.	22429.	46.32	1974.	30600.	9712.	7248.	3716215.	5350000.	1739021.	1716132.	147.7	1246.3	4009.
	6806.	22799.	45.14	1975.	31000.	9926.	6779.	3772173.	5351000.	1761044.	1603117.	161.2	1231.6	4051.
68A	7203.	23867.	42.75	1976.	31500.	10176.	6791.	3909522.	5372000.	1790290.	1661631.	170.5	1298.2	4158.
Dearborn	7585.	25030.	42.09	1977.	32200.	10495.	7303.	4071356.	5405000.	1824333.	1727408.	181.5	1369.7	4280.
	7807.	26513.	40.46	1978.	32900.	10820.	8790.	4238839.	5446000.	1861995.	1839974.	195.4	1438.6	4441.
	7950.	27317.	48.47	1979.	33700.	11185.	9011.	4319035.	5475000.	1896489.	1929580.	217.4	1479.4	4512.
	7776.	27639.	59.45	1980.	34291.	11486.	8890.	4362862.	5490180.	1927050.	1846400.	246.8	1475.0	4487.
	7649.	27954.	56.40	1981.	34800.	11765.	8572.	4374142.	5489000.	1952615.	1788511.	272.4	1512.2	4561.
	7703.	28164.	51.63	1982.	35000.	11944.	8192.	4342071.	5482000.	1976781.	1719683.	289.1	1480.0	4555.
	8960.	27885.	37.97	1970.	41546.	13692.	7889.	3136244.	5195392.	1609443.	1580100.	116.3	1085.6	3619.
	9467.	28534.	36.88	1971.	41900.	13961.	7588.	3210180.	5250000.	1645684.	1531495.	121.3	1122.4	3714.
	9749.	29662.	36.13	1972.	42300.	14252.	8980.	3439698.	5296000.	1680066.	1580411.	125.3	1185.9	3860.
	10255.	31050.	37.60	1973.	41800.	14243.	9361.	3652567.	5329000.	1711111.	1702849.	133.1	1254.3	4080.
	10193.	31937.	46.32	1974.	41500.	14302.	9291.	3716215.	5350000.	1739021.	1716132.	147.7	1246.3	4009.
	10312.	32326.	45.14	1975.	41000.	14294.	8720.	3772173.	5351000.	1761044.	1603117.	161.2	1231.6	4051.
173A	10975.	33177.	42.75	1976.	41400.	14602.	9133.	3909522.	5372000.	1790290.	1661631.	170.5	1298.2	4158.
Knox	11545.	33786.	42.09	1977.	40900.	14596.	9644.	4071356.	5405000.	1824333.	1727408.	181.5	1369.7	4280.
	12033.	34684.	40.46	1978.	41000.	14807.	12630.	4238839.	5446000.	1861995.	1839974.	195.4	1438.6	4441.
	12120.	35510.	48.47	1979.	41600.	15206.	13369.	4319035.	5475000.	1896489.	1929580.	217.4	1479.4	4512.
	11760.	36316.	59.45	1980.	41838.	15481.	13247.	4362862.	5490180.	1927050.	1846400.	246.8	1475.0	4487.
	12203.	36580.	56.40	1981.	42600.	15959.	13209.	4374142.	5489000.	1952615.	1788511.	272.4	1512.2	4561.
	12458.	36268.	51.63	1982.	42500.	16122.	13067.	4342071.	5482000.	1976781.	1719683.	289.1	1480.0	4555.

Table A2 (contd.)

	6814.	23774.	37.97	1970.	34986.	11014.	6608.	3136244.	5195392.	1609443.	1580100.	116.3	1085.6	3619.
	7030.	24513.	36.88	1971.	36200.	11507.	6865.	3210180.	5250000.	1645684.	1531495.	121.3	1122.4	3714.
	7209.	26572.	36.13	1972.	36700.	11780.	8677.	3439698.	5296000.	1680066.	1580411.	125.3	1185.9	3860.
	7553.	28396.	37.60	1973.	36900.	11962.	9371.	3652567.	5329000.	1711111.	1702849.	133.1	1254.3	4080.
	6388.	28800.	46.32	1974.	37500.	12278.	9597.	3716215.	5350000.	1739021.	1716132.	147.7	1246.3	4009.
	6514.	29550.	45.14	1975.	37800.	12502.	9087.	3772173.	5351000.	1761044.	1603117.	161.2	1231.6	4051.
254B	7232.	30825.	42.75	1976.	37900.	12663.	10109.	3909522.	5372000.	1790290.	1661631.	170.5	1298.2	4158.
Marahall	7729.	32394.	42.09	1977.	38000.	12828.	10745.	4071356.	5405000.	1824333.	1727408.	181.5	1369.7	4280.
	8351.	33271.	40.46	1978.	38600.	13166.	13288.	4238839.	5446000.	1861995.	1839974.	195.4	1438.6	4441.
	8259.	33981.	48.47	1979.	38500.	13270.	13182.	4319035.	5475000.	1896489.	1929580.	217.4	1479.4	4512.
	8092.	34240.	59.45	1980.	39155.	13640.	12200.	4362862.	5490180.	1927050.	1846400.	246.8	1475.0	4487.
	8301.	34420.	56.40	1981.	39400.	13873.	12230.	4374142.	5489000.	1952615.	1788511.	272.4	1512.2	4561.
	8459.	34247.	51.63	1982.	39700.	14131.	12126.	4342071.	5482000.	1976781.	1719683.	289.1	1480.0	4555.

a - y e - X_4 i - X_8 l - X_{11}

b - X_1 f - X_5 j - X_9 m - X_{12}

c - X_2 g - X_6 k - X_{10} n - X_{13}

d - X_3 h - X_7

Note: For the meaning and definition of each variable, see Table 4.1 and Chapter 4 in the text.

Table A3

Data Table for Rural Minor Arterial

Station County	a	b	c	d	e	f	g
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	3245.	21544.	37.97	1970.	31382.	9696.	7710.
	3276.	21837.	36.88	1971.	31400.	9791.	7198.
	3478.	23536.	36.13	1972.	32300.	10167.	8165.
	3977.	25371.	37.60	1973.	33000.	10486.	8351.
	3843.	25356.	46.32	1974.	33500.	10747.	9138.
	3739.	25653.	45.14	1975.	33500.	10851.	7545.
25A	3913.	26833.	42.75	1976.	33800.	11056.	8061.
Noble	4071.	28220.	42.09	1977.	34100.	11264.	9432.
	4251.	29281.	40.46	1978.	34600.	11543.	9894.
	4051.	29889.	48.47	1979.	35400.	11929.	10029.
	3848.	29979.	59.45	1980.	35443.	12065.	9244.
	3885.	29891.	56.40	1981.	35000.	12037.	9426.
	3898.	29501.	51.63	1982.	35300.	12266.	9129.
	3298.	14036.	37.97	1970.	21138.	6454.	4261.
	3545.	14392.	36.88	1971.	21700.	6687.	4279.
	3620.	15165.	36.13	1972.	22000.	6843.	4756.
	3634.	16070.	37.60	1973.	22400.	7033.	5063.
	3554.	16461.	46.32	1974.	22900.	7258.	5263.
	3624.	16959.	45.14	1975.	23500.	7520.	5398.
301A	3840.	17679.	42.75	1976.	23700.	7658.	5554.
Ripley	3920.	18397.	42.09	1977.	24000.	7831.	5694.
	4049.	18973.	40.46	1978.	24100.	7941.	7027.
	4119.	19503.	48.47	1979.	24500.	8154.	7466.
	3845.	19869.	59.45	1980.	24398.	8202.	7287.
	3798.	20223.	56.40	1981.	24600.	8354.	7024.
	3740.	20112.	51.63	1982.	24700.	8475.	7048.
	5572.	26922.	37.97	1970.	44176.	12900.	3693.
	5697.	28184.	36.88	1971.	44500.	13146.	3749.
	6049.	30834.	36.13	1972.	44500.	13300.	4440.
	5850.	33276.	37.60	1973.	46600.	14095.	4732.
	5913.	34425.	46.32	1974.	47300.	14479.	4921.
	5970.	35729.	45.14	1975.	48100.	14904.	4939.
313A	6079.	37434.	42.75	1976.	48500.	15214.	5433.
Morgan	5895.	39589.	42.09	1977.	49800.	15817.	6010.
	5980.	41534.	40.46	1978.	50600.	16275.	8286.
	6010.	42775.	48.47	1979.	51200.	16680.	8612.
	5650.	43414.	59.45	1980.	51999.	17160.	8430.
	5631.	43802.	56.40	1981.	52500.	17554.	8271.
	5565.	43702.	51.63	1982.	52600.	17822.	8317.

Table A3 (continued)

Station County	a	b	c	d	e	f	g
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	2571.	15947.	37.97	1970.	20995.	6872.	3806.
	2439.	16164.	36.88	1971.	21300.	7051.	3873.
	2452.	17082.	36.13	1972.	21300.	7132.	4856.
	2512.	18200.	37.60	1973.	21900.	7417.	5322.
	2389.	18756.	46.32	1974.	22000.	7539.	5340.
	2415.	19165.	45.14	1975.	22100.	7663.	5092.
262A	2564.	20024.	42.75	1976.	22600.	7931.	5672.
White	2555.	20852.	42.09	1977.	22900.	8134.	5609.
	2687.	21497.	40.46	1978.	23300.	8378.	7071.
	2527.	22007.	48.47	1979.	23400.	8518.	7637.
	2444.	22321.	59.45	1980.	23867.	8798.	7247.
	2460.	22634.	56.40	1981.	23800.	8885.	7305.
	2436.	22579.	51.63	1982.	24000.	9076.	6862.

a = y

b = X_1

c = X_2

d = X_3

e = X_4

f = X_5

g = X_6

Note: For the meaning and definition of each variable, see Table 4.1 and Chapter 4 in the text.

Table A4

Data Table for Rural Major Collector

Station County	a	b	c	d	e	f	g
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1064.	19885.	37.97	1970.	28915.	9645.	6884.
	1115.	20226.	36.88	1971.	29500.	9905.	6876.
	1236.	21122.	36.13	1972.	29600.	10005.	7852.
	1118.	22220.	37.60	1973.	29900.	10174.	8495.
	1117.	22770.	46.32	1974.	29800.	10208.	8593.
	1093.	23153.	45.14	1975.	29800.	10278.	7517.
47A	1064.	23697.	42.75	1976.	29800.	10348.	7544.
Randolph	1057.	25175.	42.09	1977.	30200.	10559.	8555.
	1159.	25362.	40.46	1978.	29800.	10491.	10037.
	1161.	25785.	48.47	1979.	30100.	10671.	9634.
	1132.	25580.	51.63	1982.	29000.	10501.	7848.
	3667.	23174.	37.97	1970.	35096.	10792.	3676.
	3850.	24314.	36.88	1971.	35200.	10896.	3935.
	4083.	26423.	36.13	1972.	36700.	11437.	4756.
	4250.	28570.	37.60	1973.	38400.	12048.	5359.
	4290.	29951.	46.32	1974.	39600.	12509.	5480.
	4391.	31284.	45.14	1975.	40100.	12754.	5256.
59A	4634.	32485.	42.75	1976.	40700.	13034.	5588.
Hancock	4420.	35745.	42.09	1977.	41400.	13351.	5889.
	4707.	37448.	40.46	1978.	42100.	13672.	7981.
	4707.	38695.	48.47	1979.	43200.	14128.	8397.
	4660.	38017.	59.45	1980.	43939.	14472.	8432.
	4346.	38234.	56.40	1981.	43900.	14563.	8038.
	4368.	38447.	51.63	1982.	43800.	14634.	7769.
	1722.	22561.	37.97	1970.	33930.	11044.	7516.
	1719.	23115.	36.88	1971.	34300.	11287.	7368.
	1845.	24694.	36.13	1972.	34600.	11513.	8644.
	1859.	26075.	37.60	1973.	35000.	11777.	9009.
	1772.	26619.	46.32	1974.	35200.	11979.	9308.
	1815.	27266.	45.14	1975.	35400.	12186.	9195.
5420A	1863.	28460.	42.75	1976.	35400.	12328.	9667.
Montgo-	1901.	28696.	42.09	1977.	35500.	12508.	9858.
mery	2248.	29514.	40.46	1978.	35600.	12693.	12251.
	2586.	29760.	48.47	1979.	35600.	12846.	12434.
	2139.	30356.	59.45	1980.	35501.	12967.	11852.
	1970.	30412.	56.40	1981.	34900.	12905.	11663.
	1890.	30148.	51.63	1982.	35300.	13216.	11313.

a = y b = X_1 c = X_2 d = X_3 e = X_4 f = X_5 g = X_6

Note: For the meaning and definition of each variable, see Table 4.1 and Chapter 4 in the text.

Appendix B

Scatter Plots:

Aggregate Analysis

1. Rural Interstate: Figure B1.1 to Figure B1.13
2. Rural Principal Arterial: Figure B2.1 to Figure B2.13
3. Rural Minor Arterial: Figure B3.1 to Figure B3.6
4. Rural Major Collector: Figure B4.1 to Figure B4.6

FIGURE B1.1: AADT VS. COUNTY VEHICLE REGISTRATION

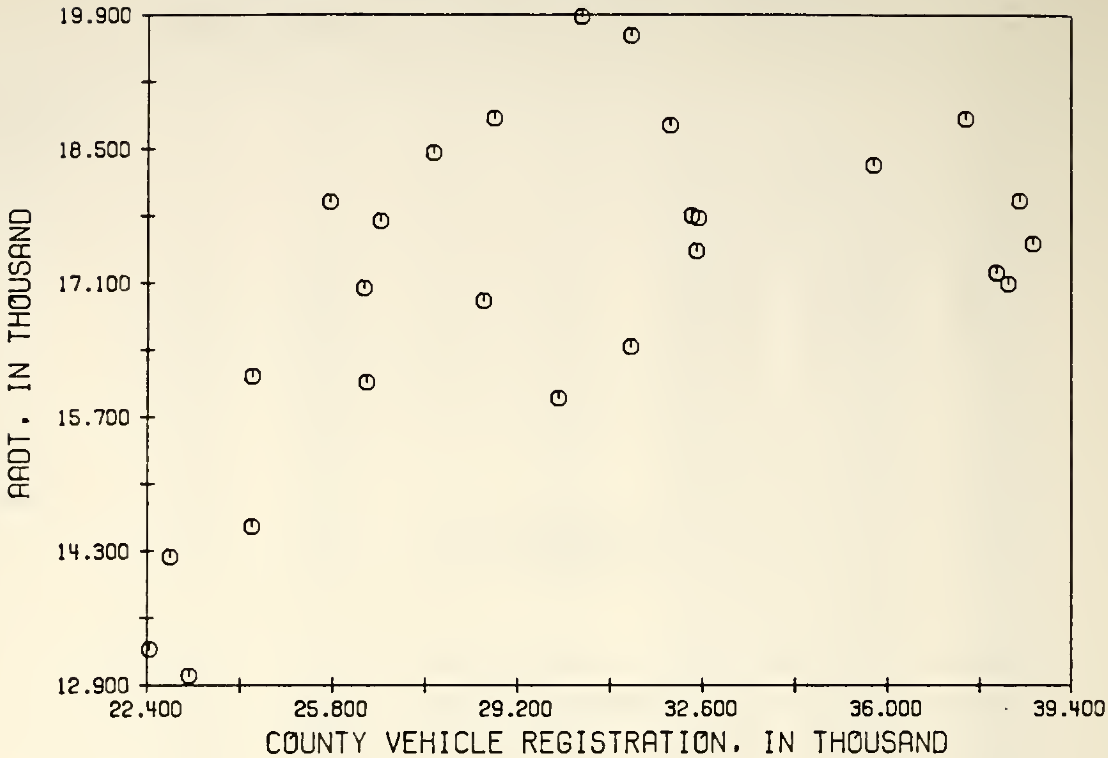


FIGURE B1.2: AADT VS. US GASOLINE PRICE

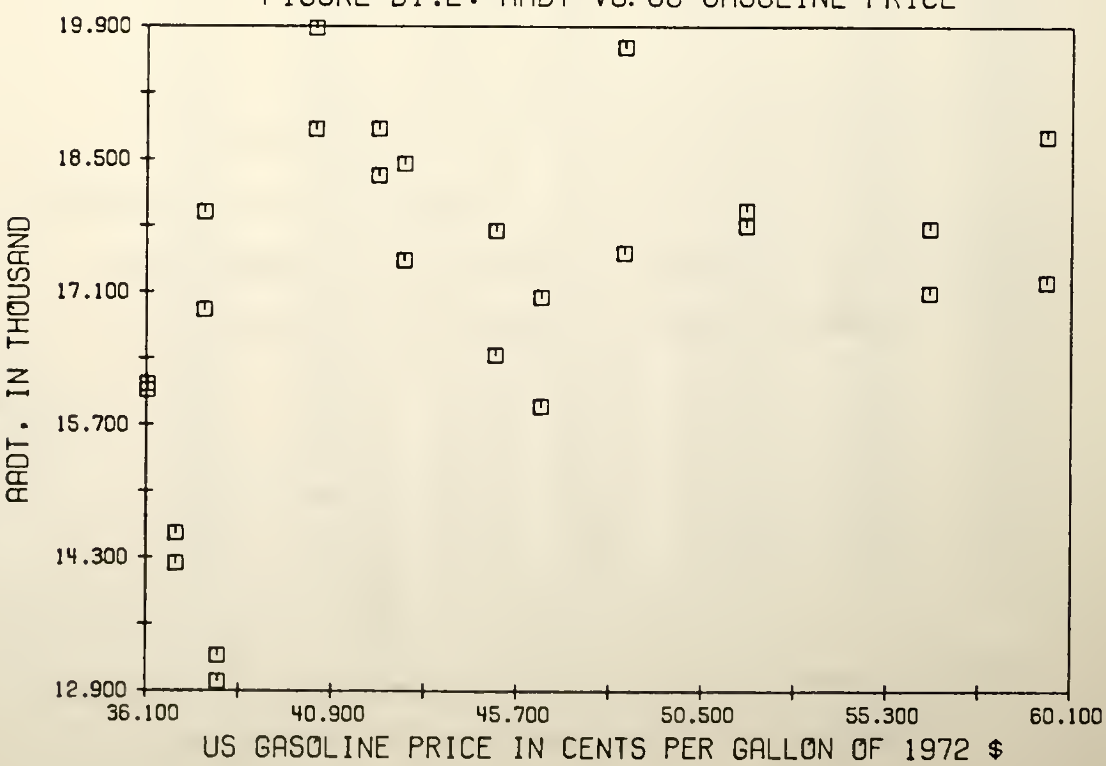


FIGURE B1.3: AADT VS. YEAR

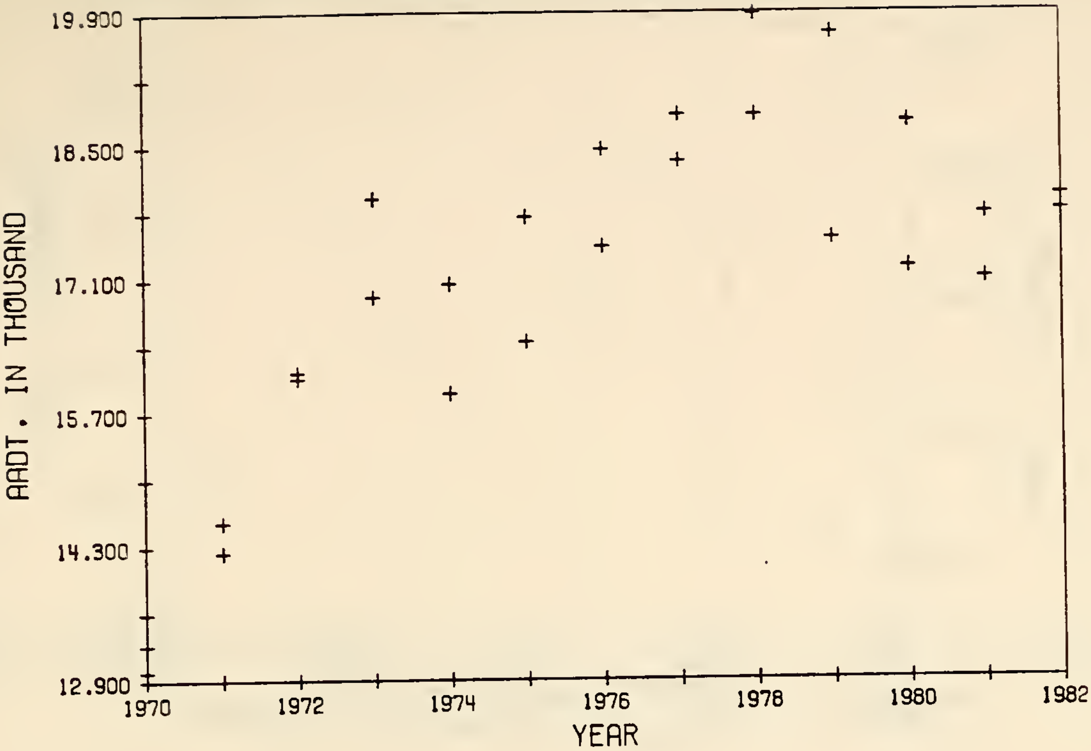


FIGURE B1.4: AADT VS. COUNTY POPULATION

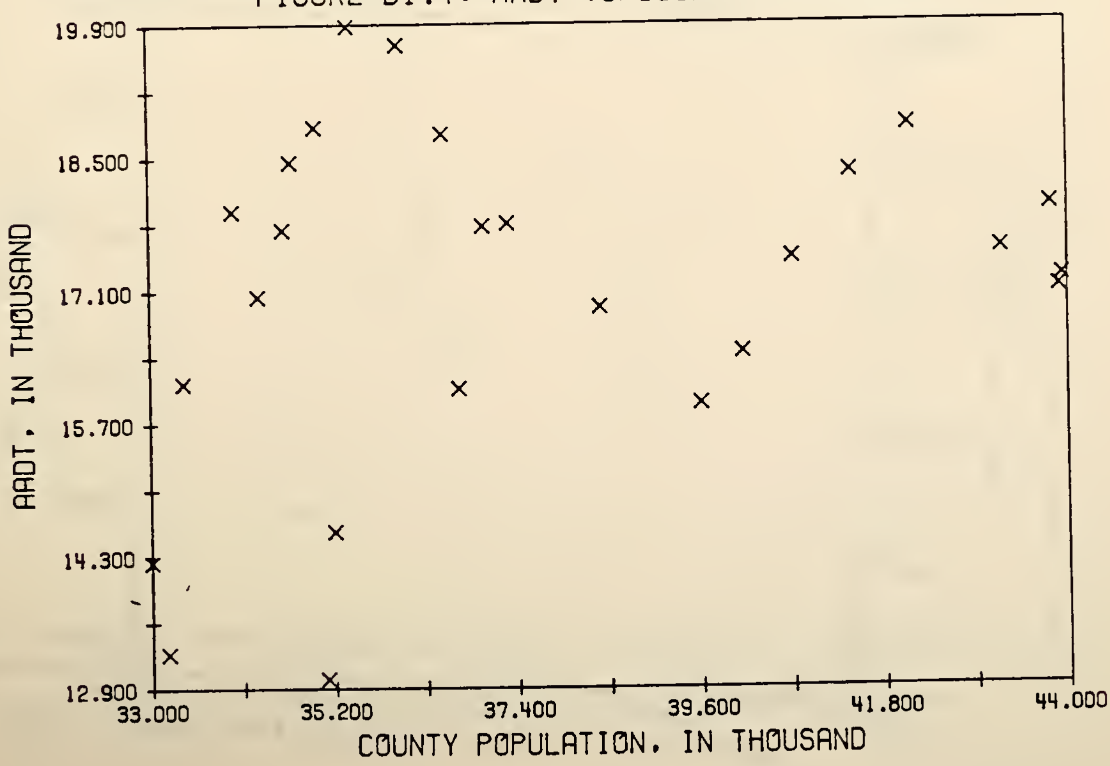


FIGURE B1.5: AADT VS. COUNTY HOUSEHOLDS

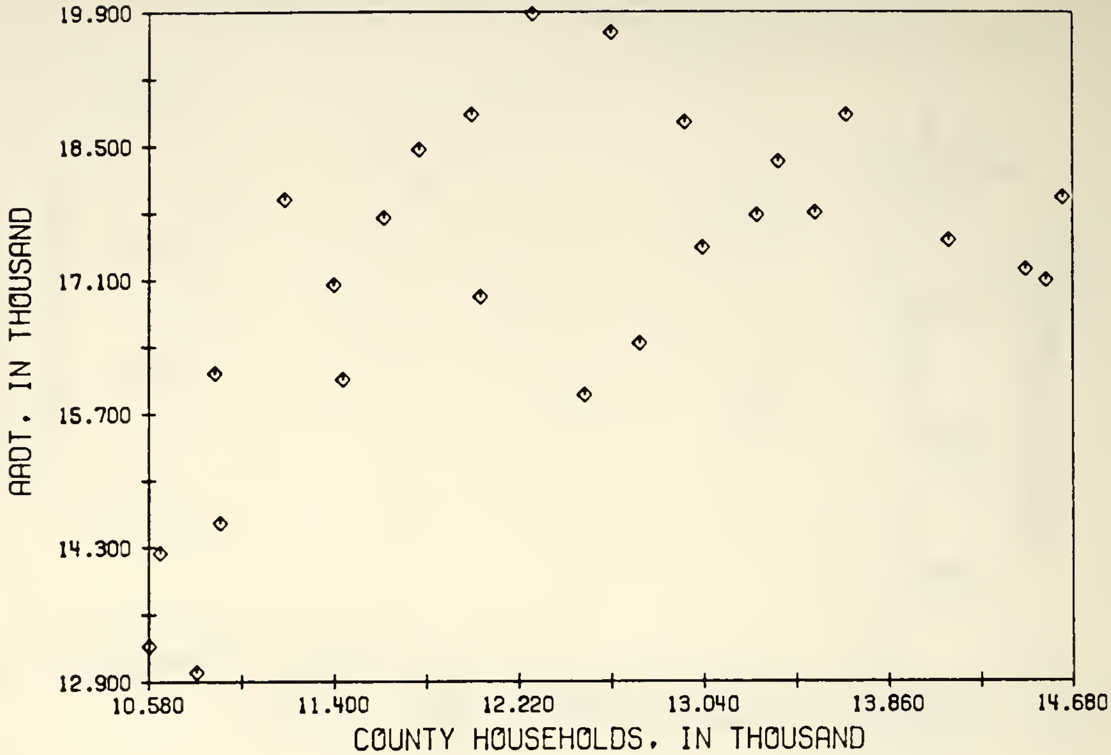


FIGURE B1.6: AADT VS. COUNTY EMPLOYMENT

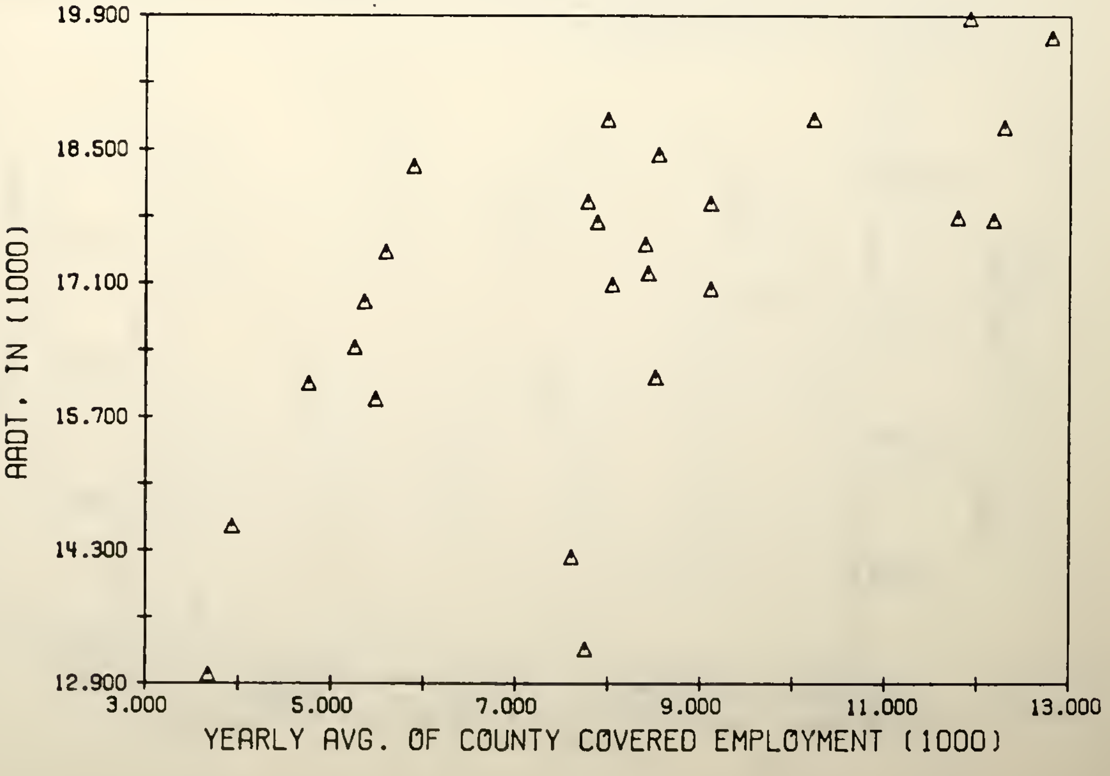


FIGURE B1.7: AADT VS. STATE VEHICLE REGISTRATION

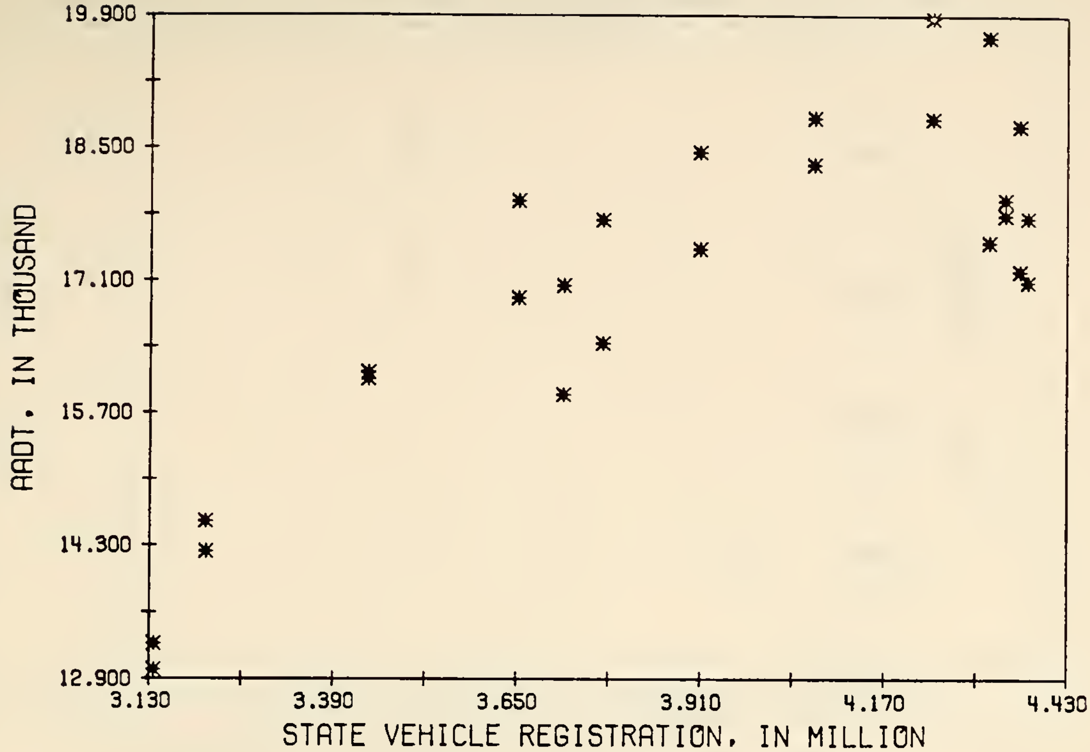


FIGURE B1.8: AADT VS. STATE POPULATION

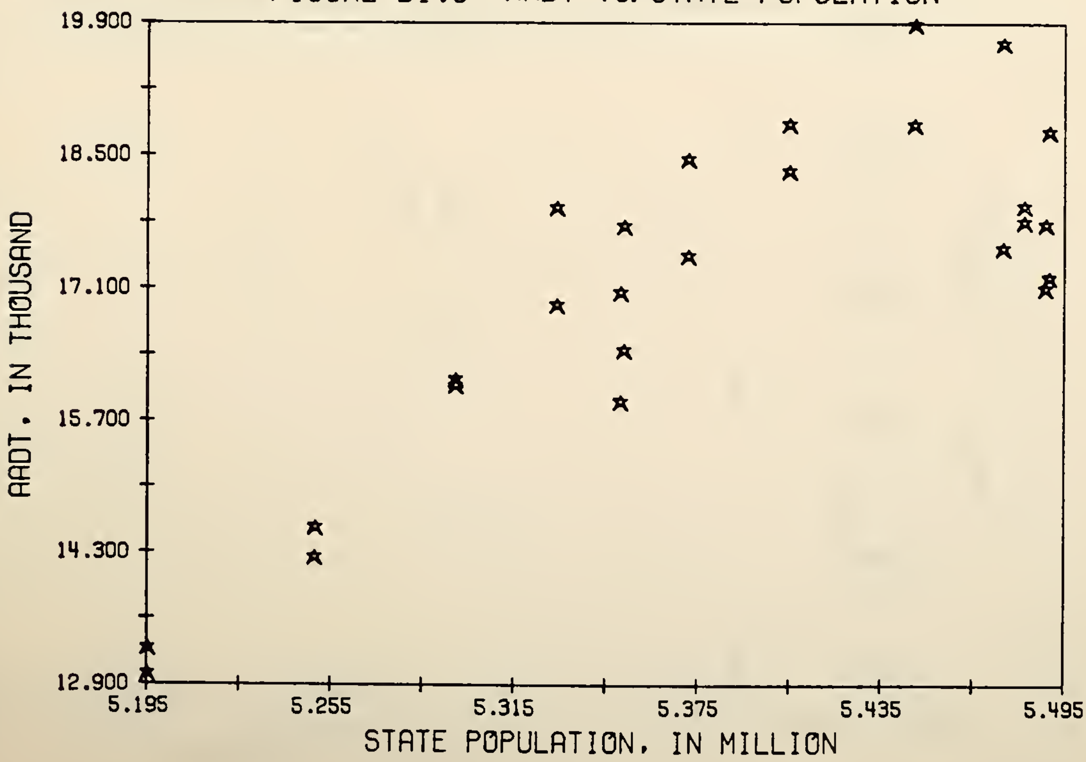


FIGURE B1.9: AADT VS. STATE HOUSEHOLDS

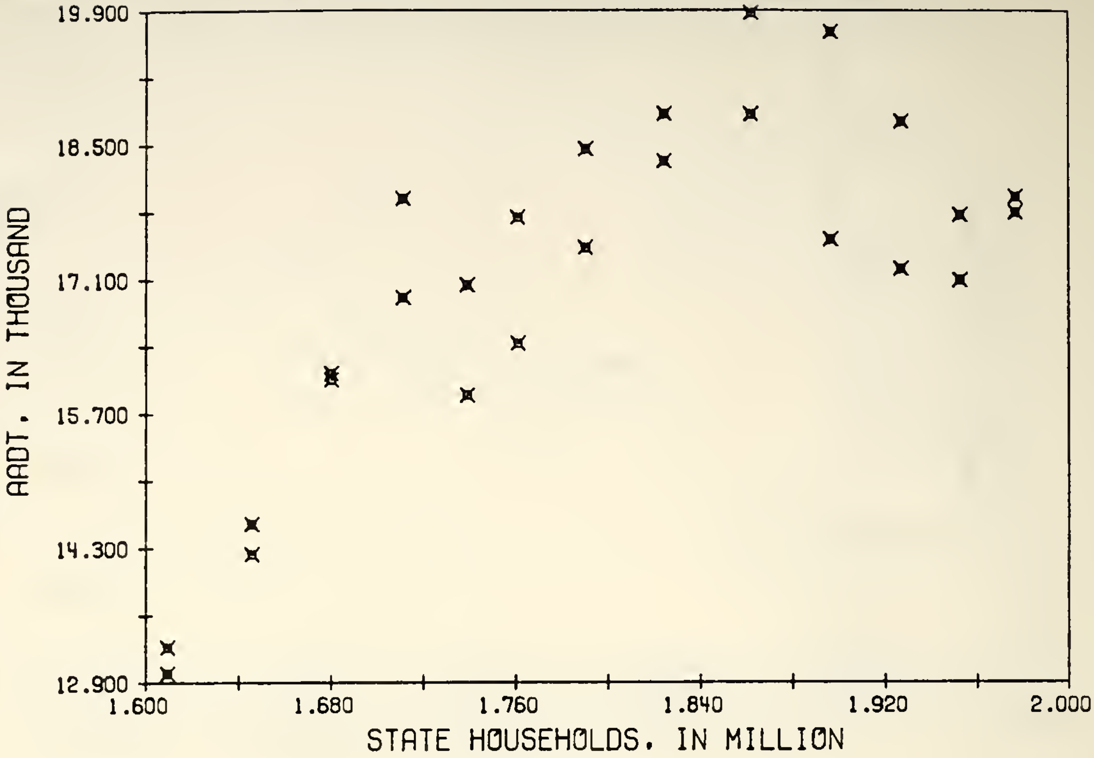


FIGURE B1.10: AADT VS. STATE EMPLOYMENT

FIGURE B1.11: AADT VS. CONSUMER PRICE INDEX

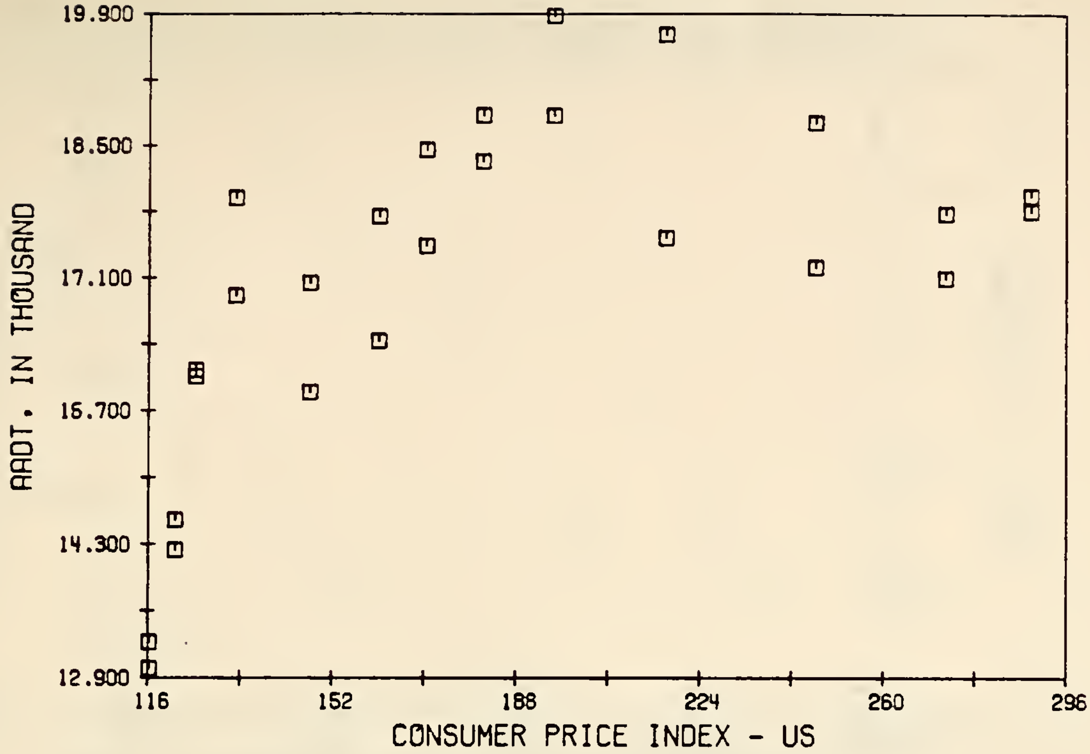


FIGURE B1.12: AADT VS. GROSS NATIONAL PRODUCT

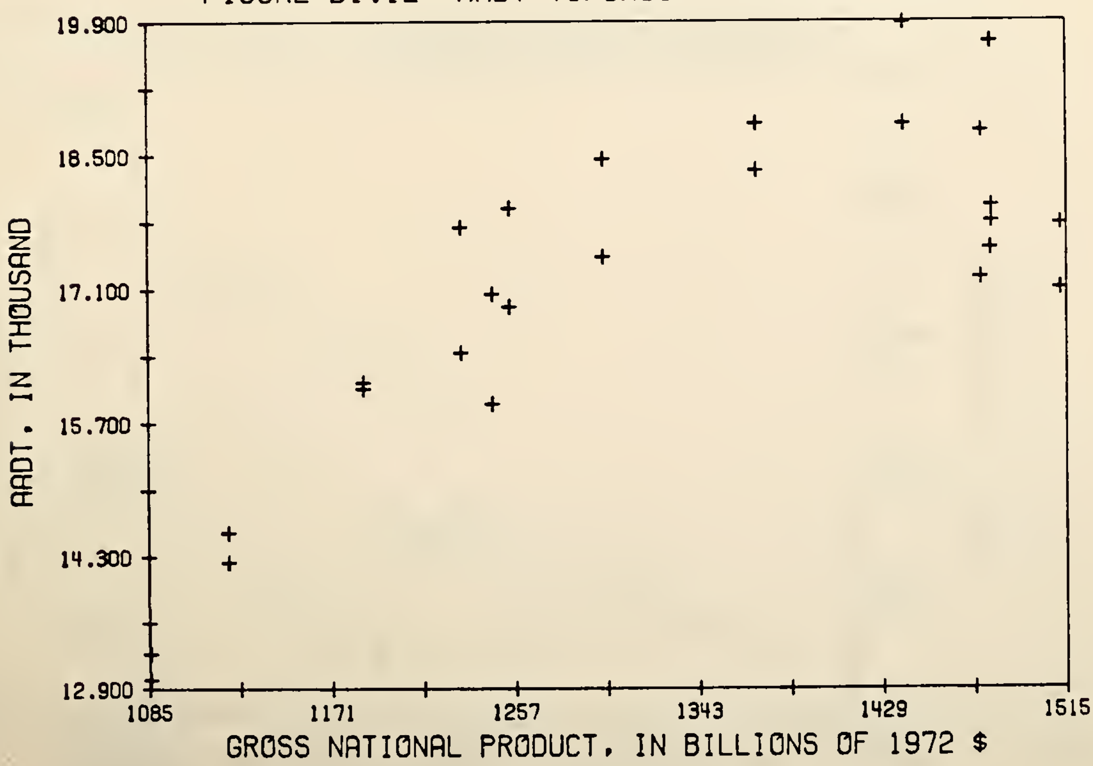


FIGURE B1.13: AADT VS. PER CAPITA NATIONAL INCOME

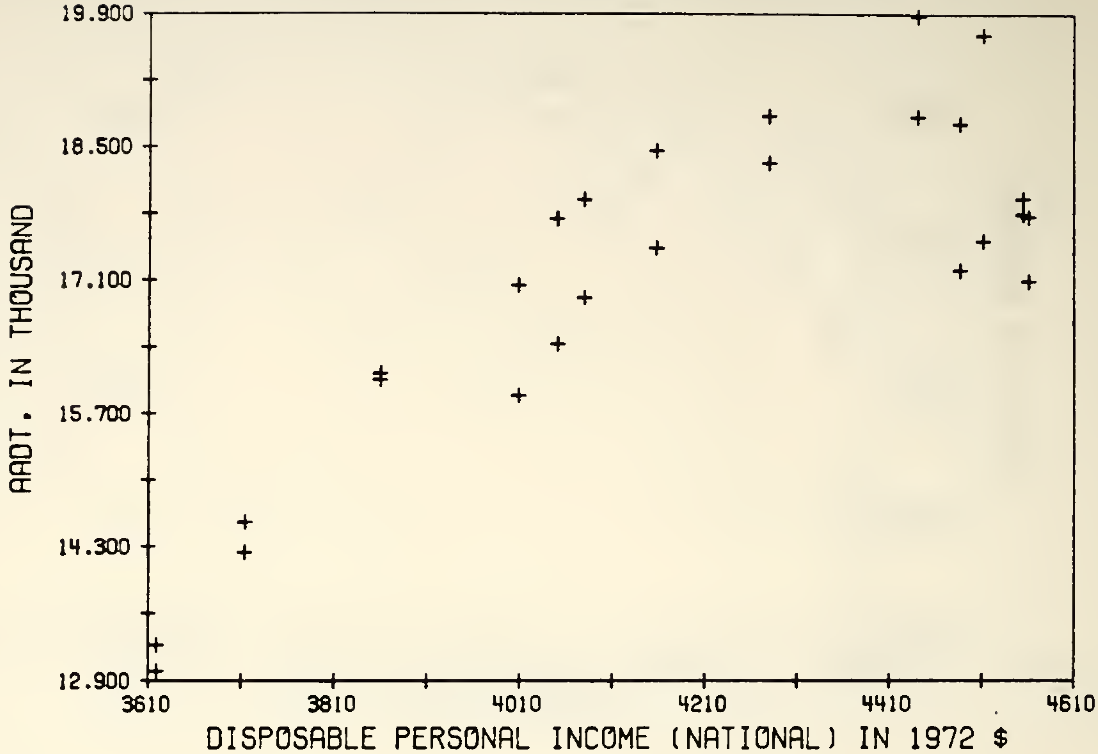


FIGURE B2.1: AADT VS. COUNTY VEHICLE REGISTRATION

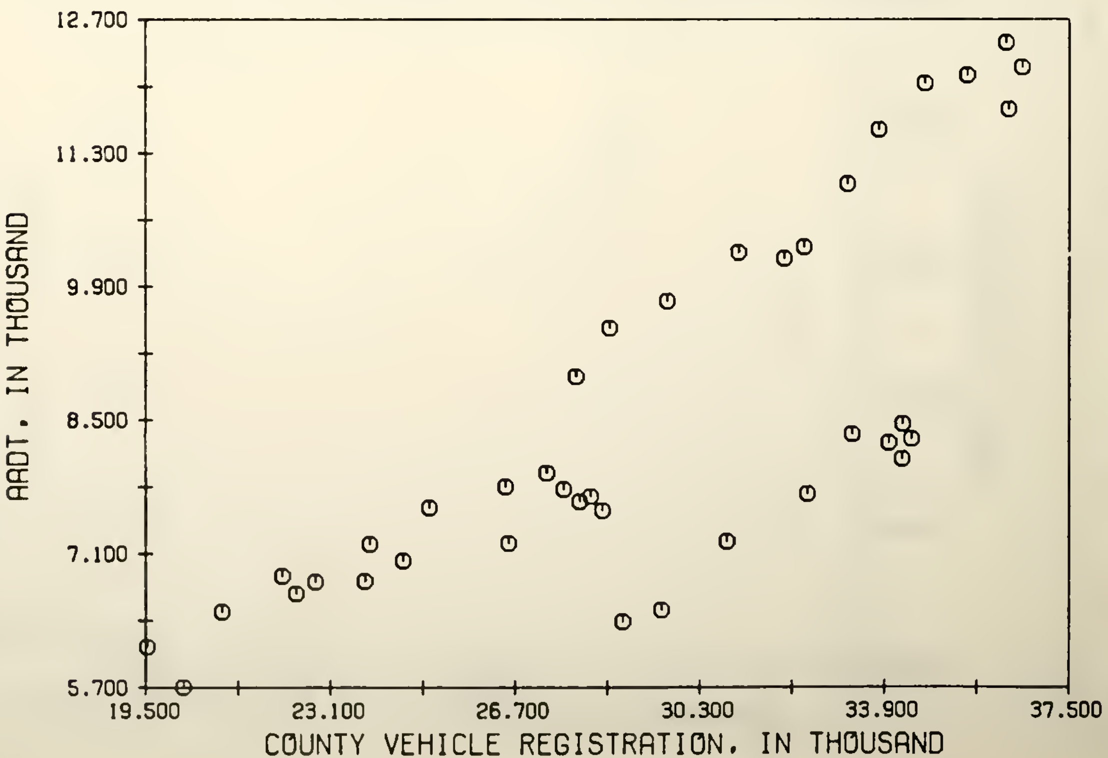


FIGURE B2.2: AADT VS. US GASOLINE PRICE

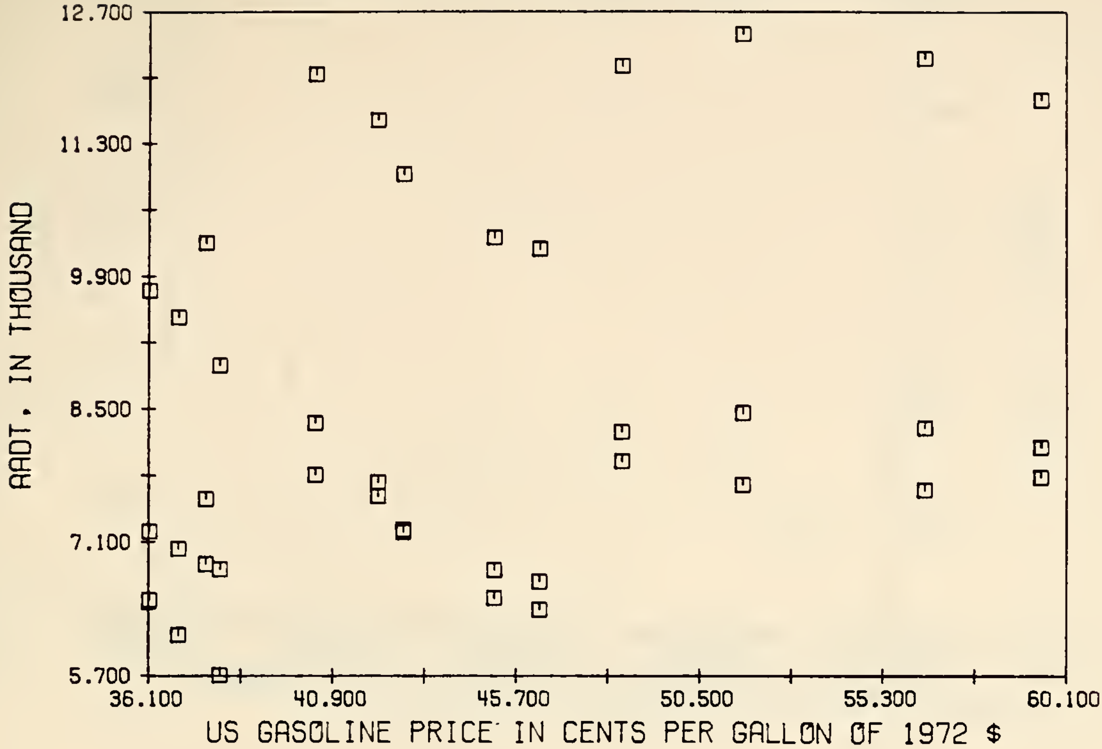


FIGURE B2.3: AADT VS. YEAR

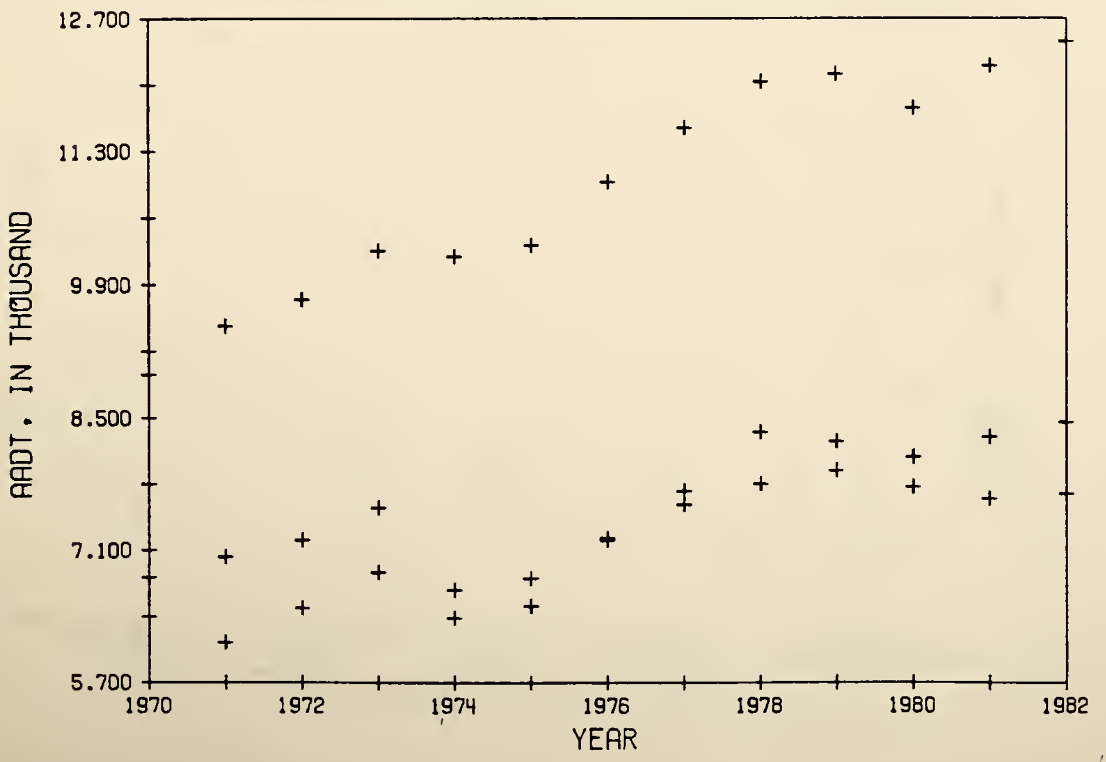


FIGURE B2.4: AADT VS. COUNTY POPULATION

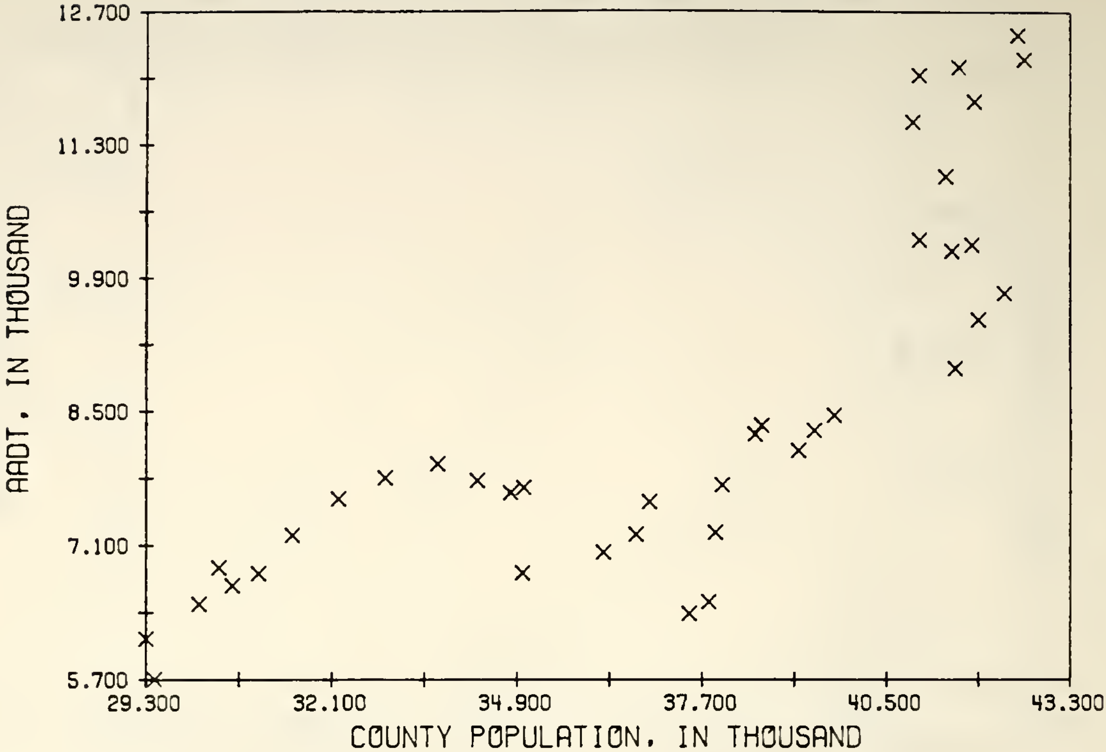


FIGURE B2.5: AADT VS. COUNTY HOUSEHOLDS

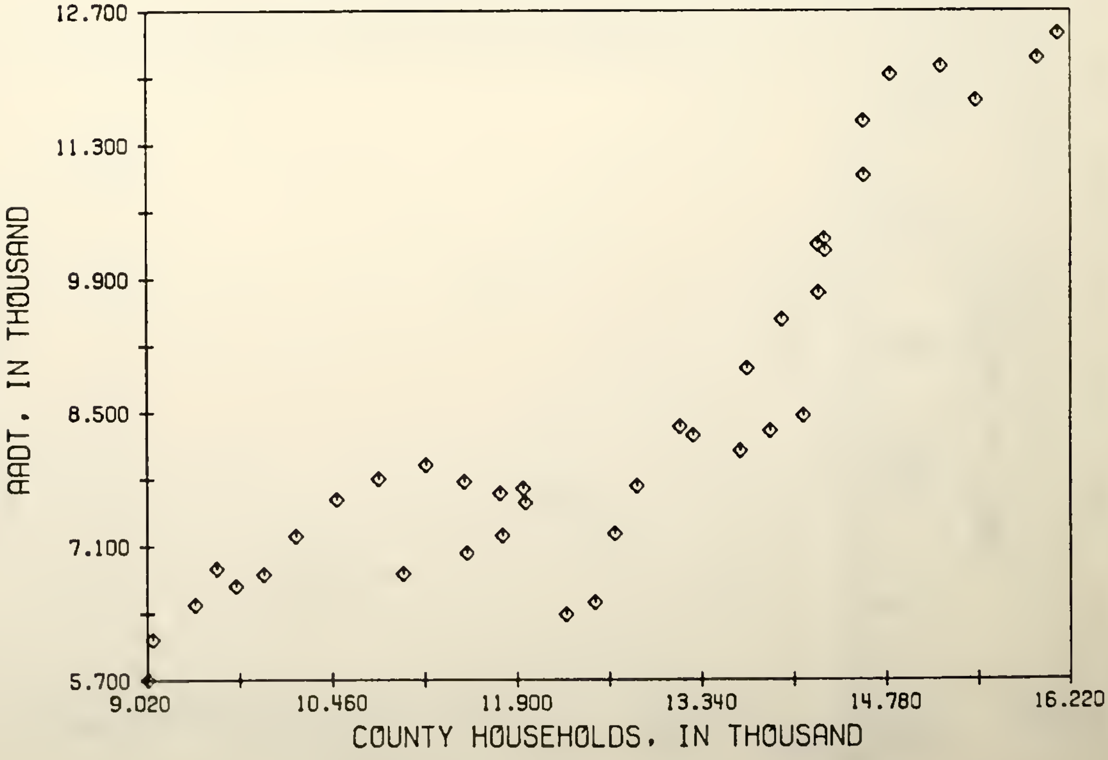


FIGURE B2.6: AADT VS. COUNTY EMPLOYMENT

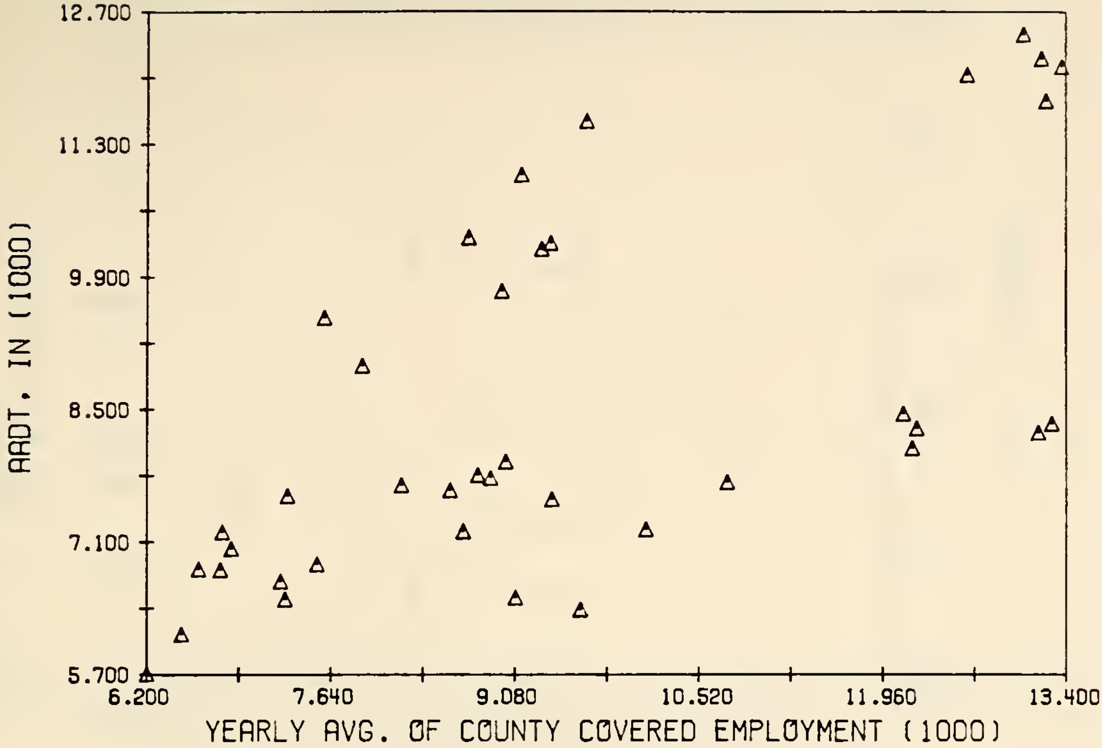


FIGURE B2.7: AADT VS. STATE VEHICLE REGISTRATION

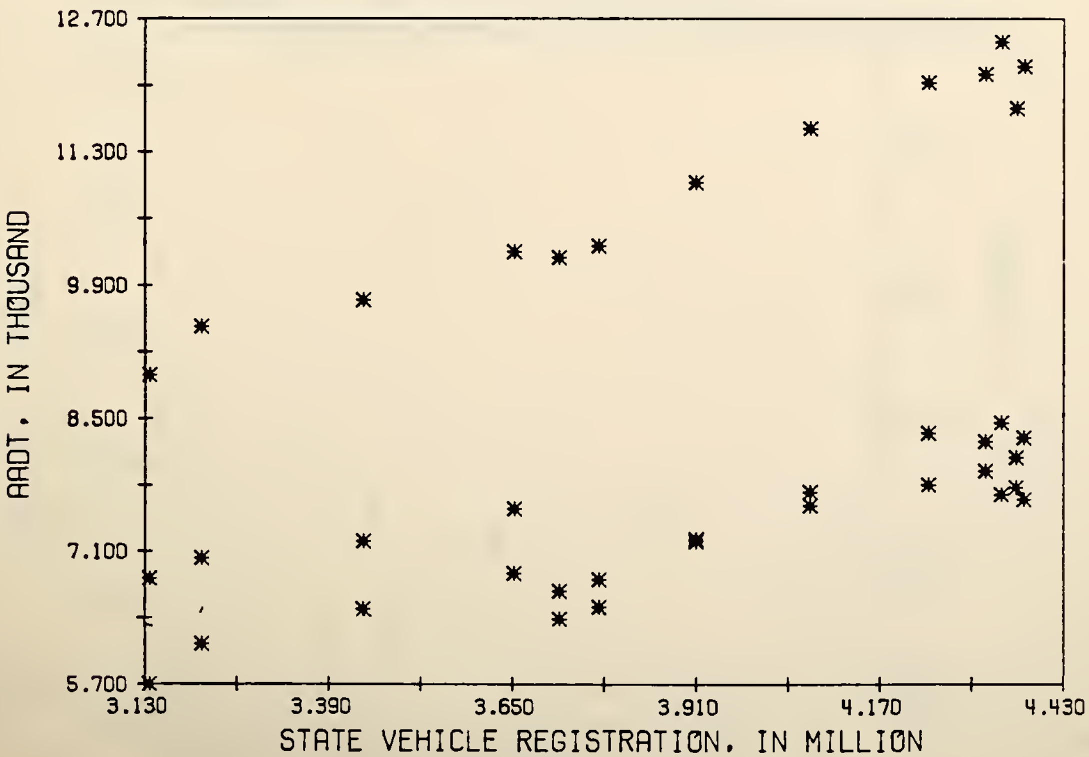


FIGURE B2.8: AADT VS. STATE POPULATION

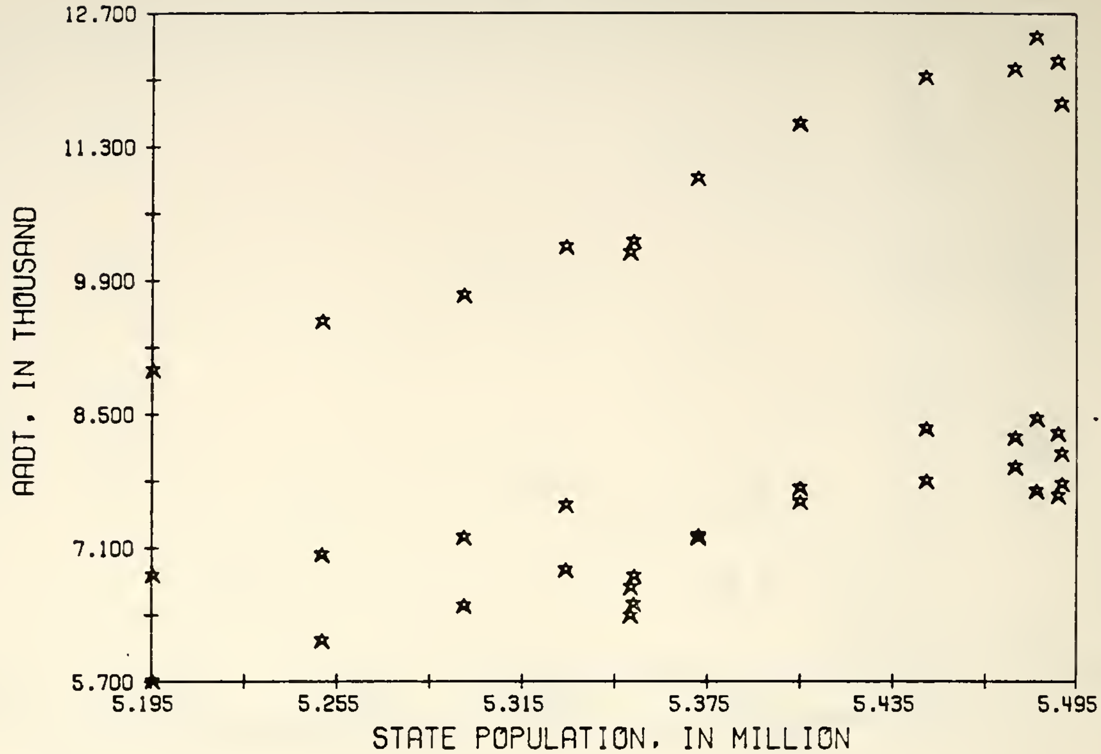


FIGURE B2.9: AADT VS. STATE HOUSEHOLDS

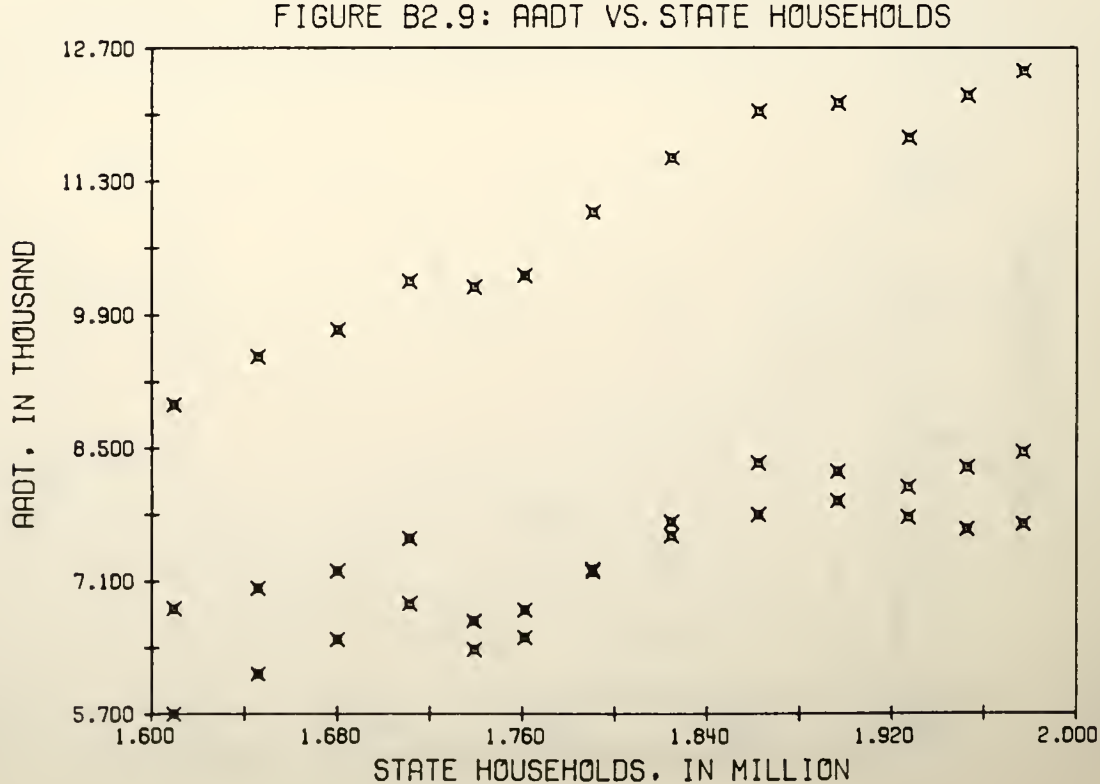


FIGURE B2.10: AADT VS. STATE EMPLOYMENT

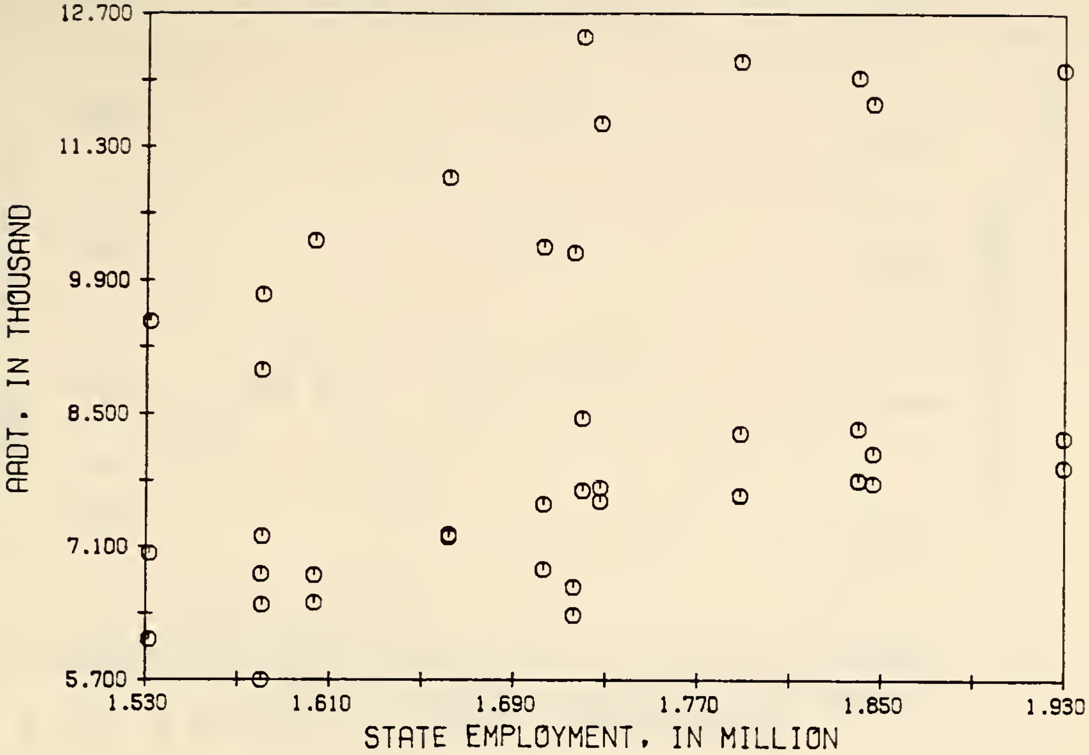


FIGURE B2.11: AADT VS. CONSUMER PRICE INDEX

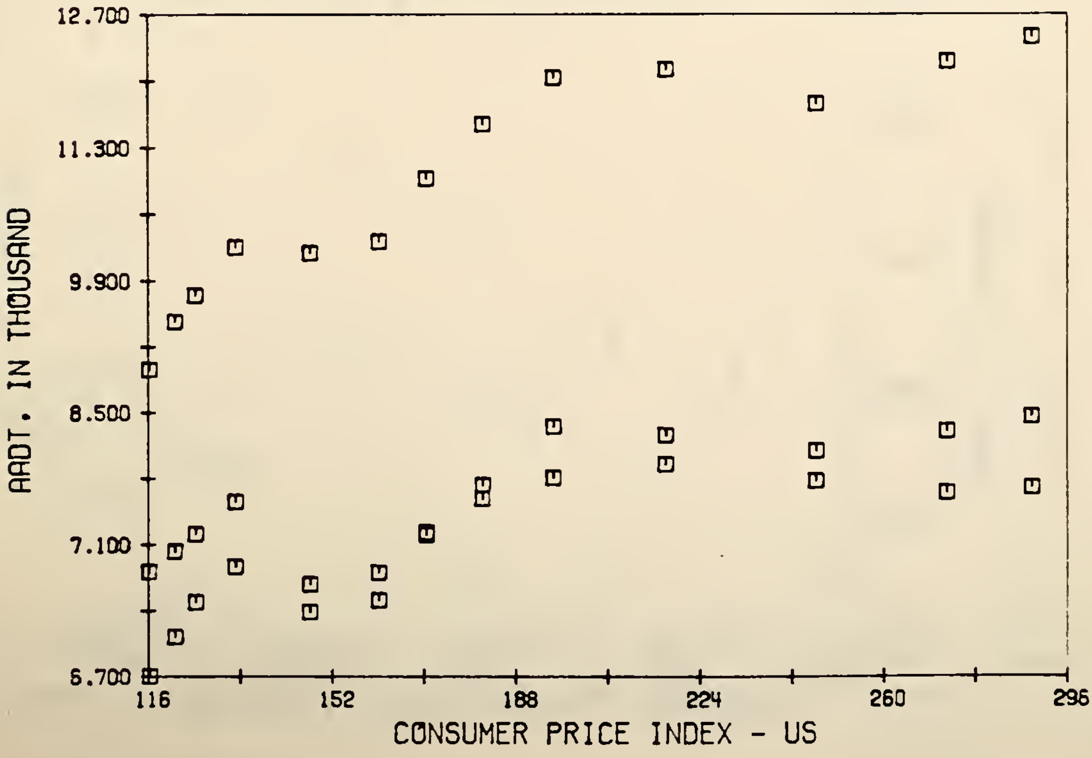


FIGURE B2.12: AADT VS. GROSS NATIONAL PRODUCT

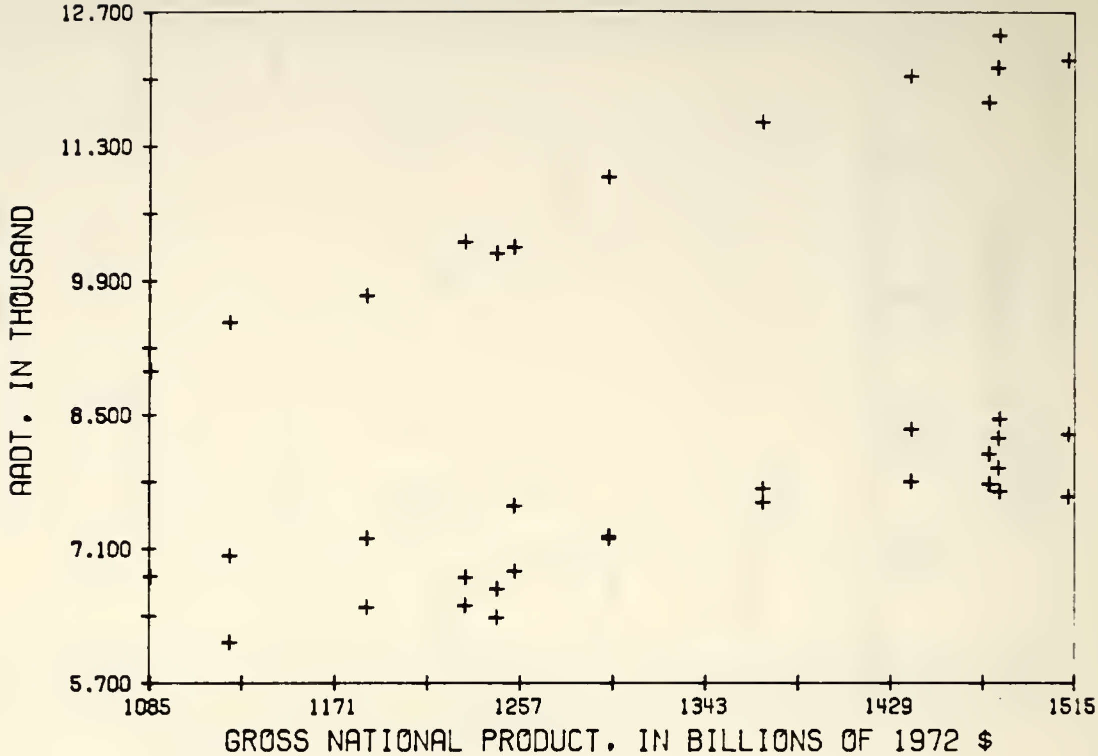


FIGURE B2.13: AADT VS. PER CAPITA NATIONAL INCOME

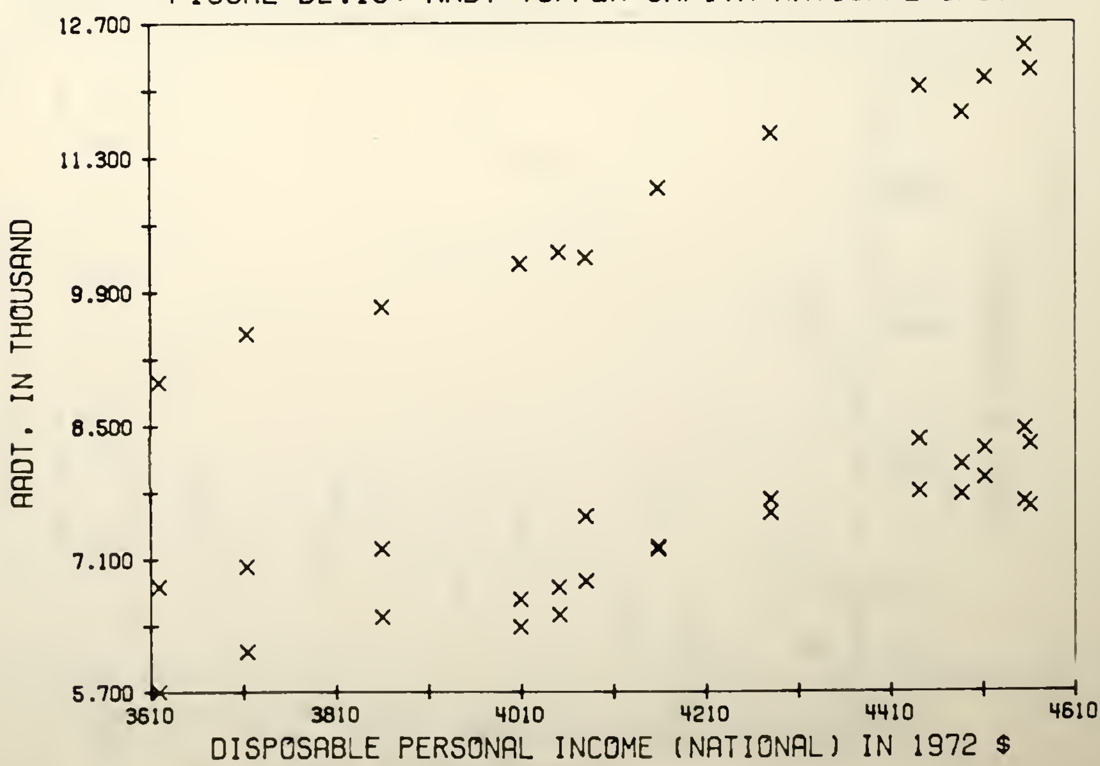


FIGURE B3.1: AADT VS. COUNTY VEHICLE REGISTRATION

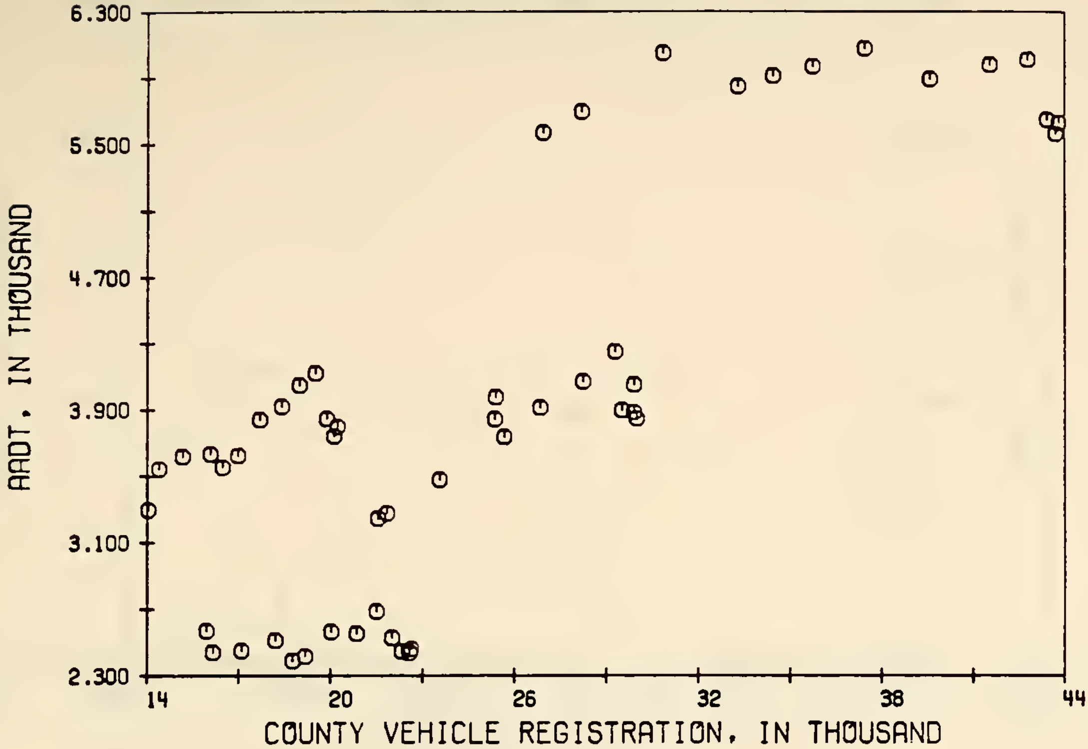


FIGURE B3.2: AADT VS. US GASOLINE PRICE

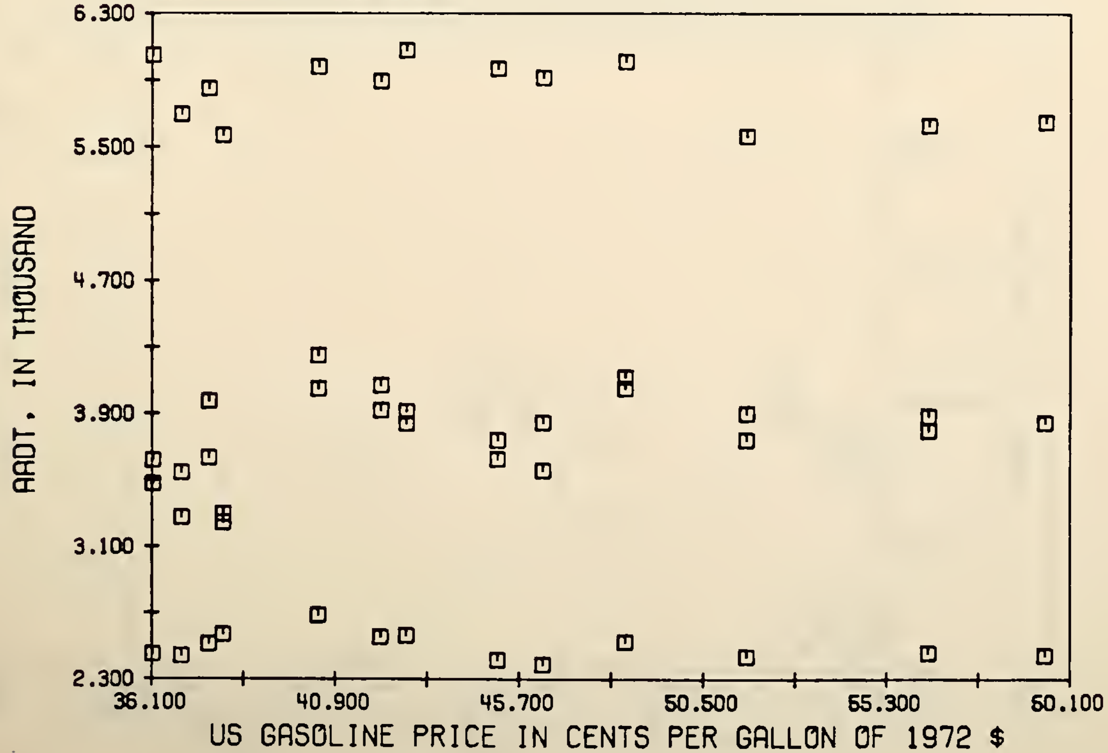


FIGURE B3.3: AADT VS. YEAR

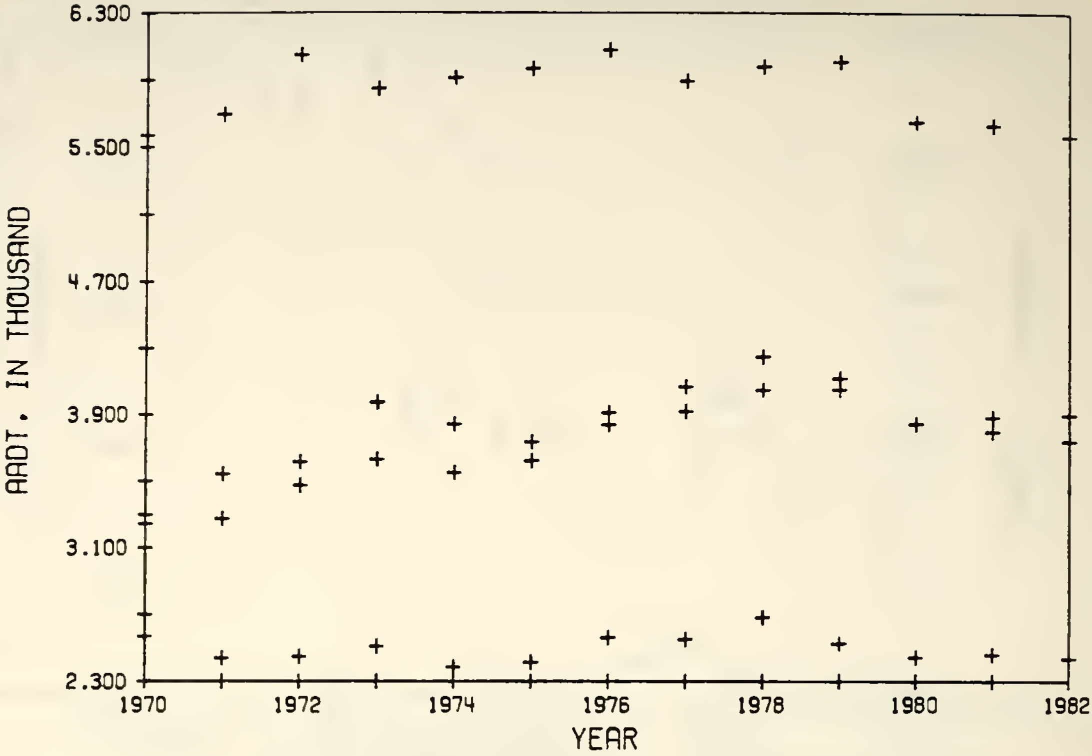


FIGURE B3.4: AADT VS. COUNTY POPULATION

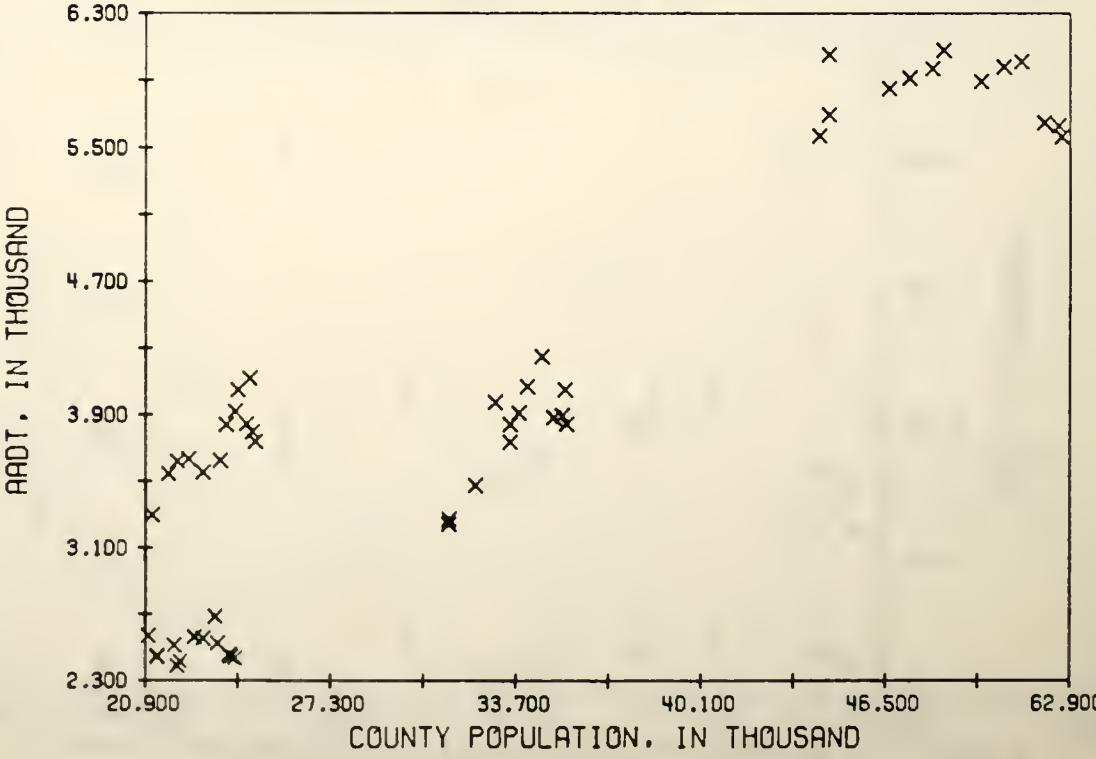


FIGURE B3.5: AADT VS. COUNTY HOUSEHOLDS

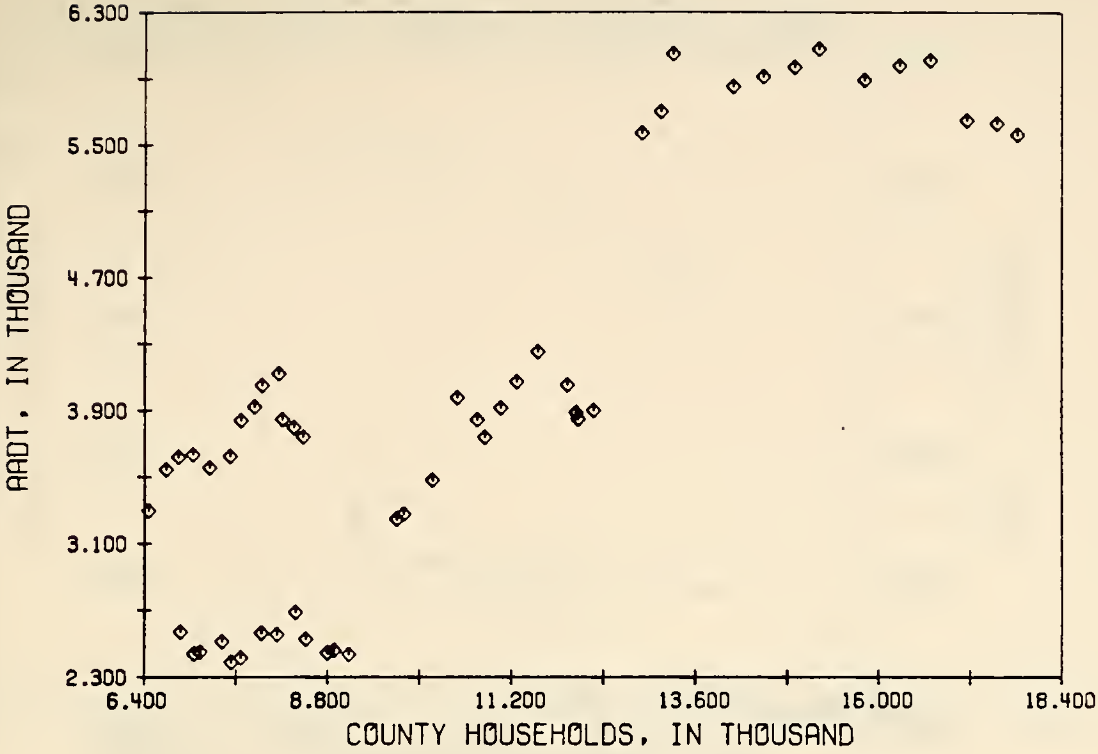


FIGURE B3.6: AADT VS. COUNTY EMPLOYMENT

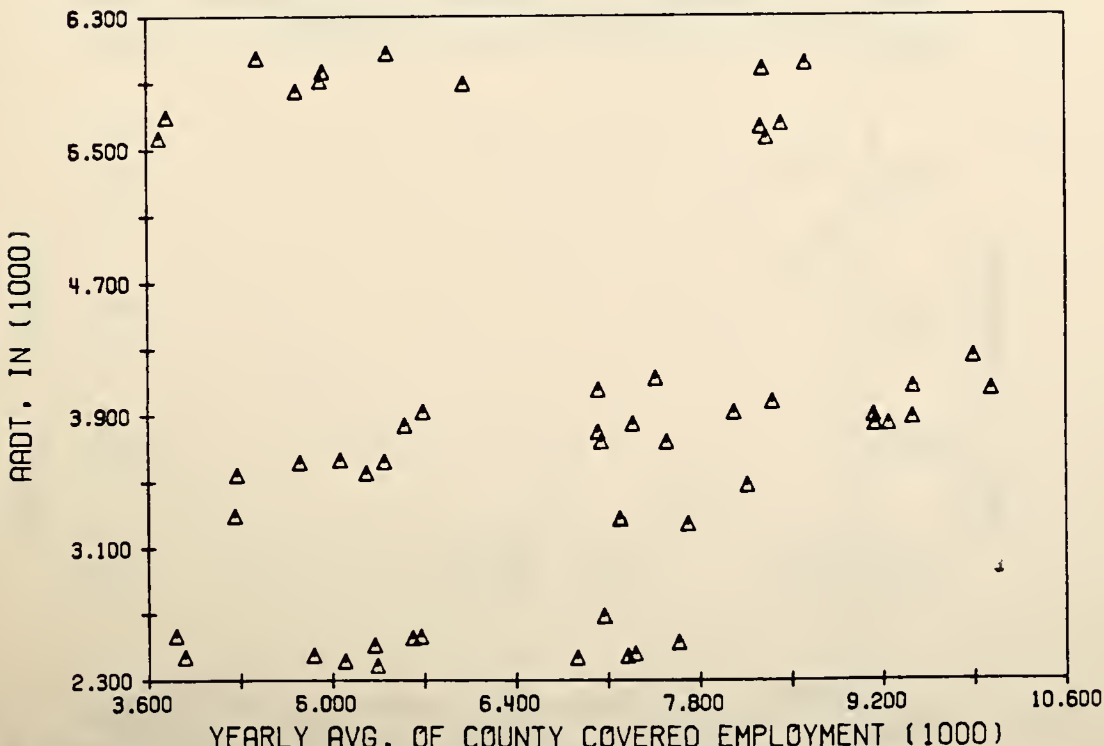


FIGURE B4.1: AADT VS. COUNTY VEHICLE REGISTRATION

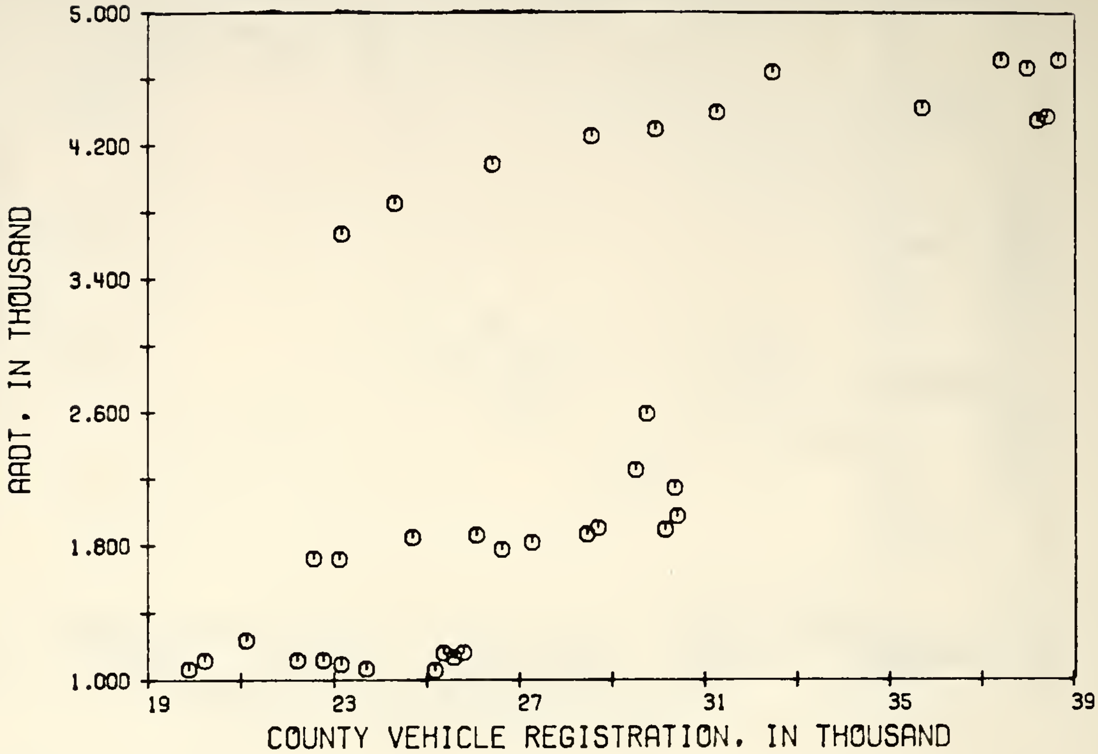


FIGURE B4.2: AADT VS. US GASOLINE PRICE

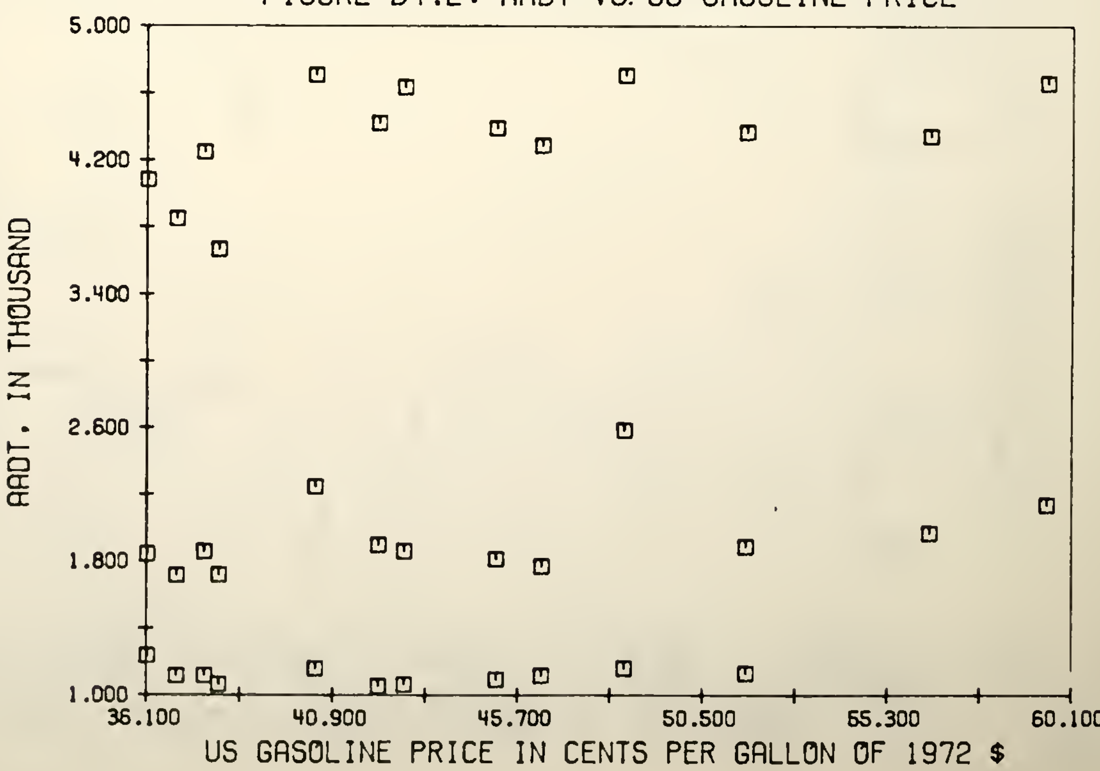


FIGURE B4.3: AADT VS. YEAR

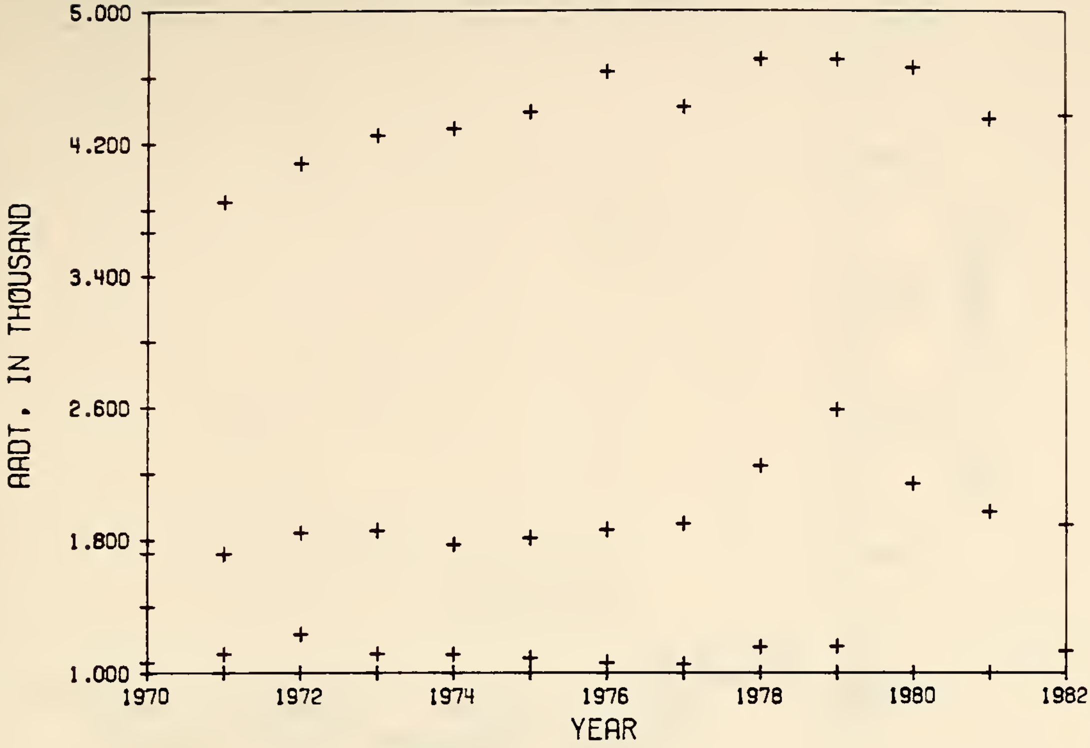


FIGURE B4.4: AADT VS. COUNTY POPULATION

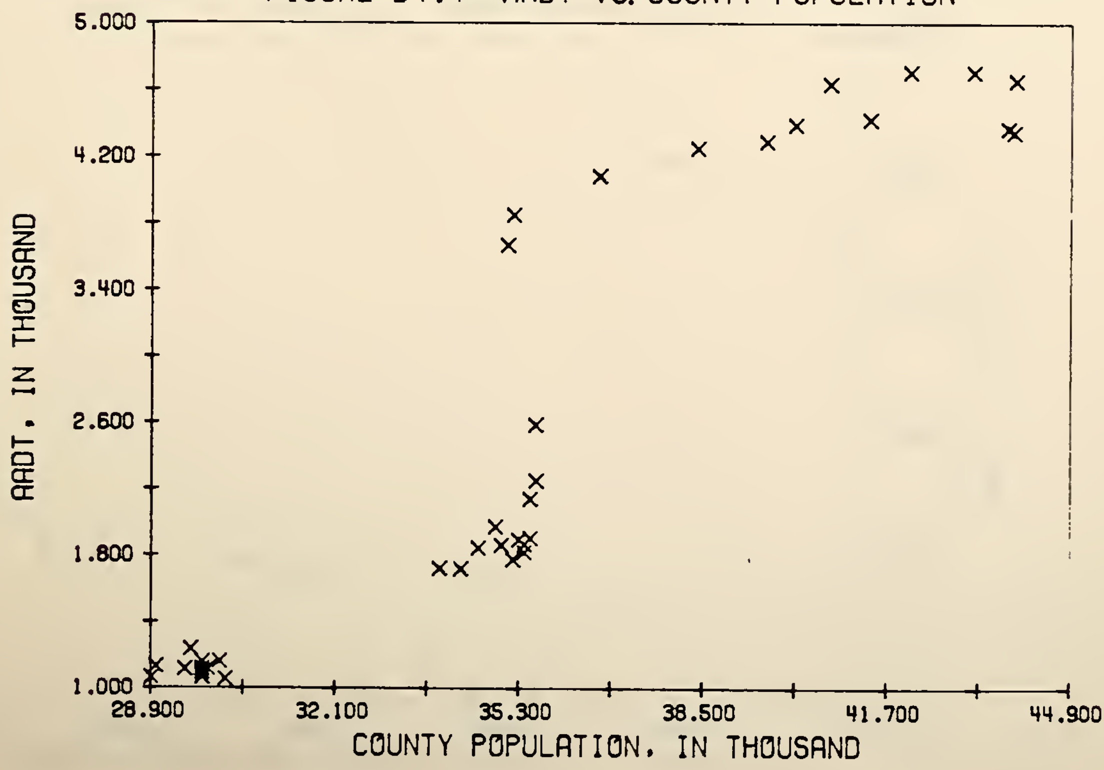


FIGURE B4.5: AADT VS. COUNTY HOUSEHOLDS

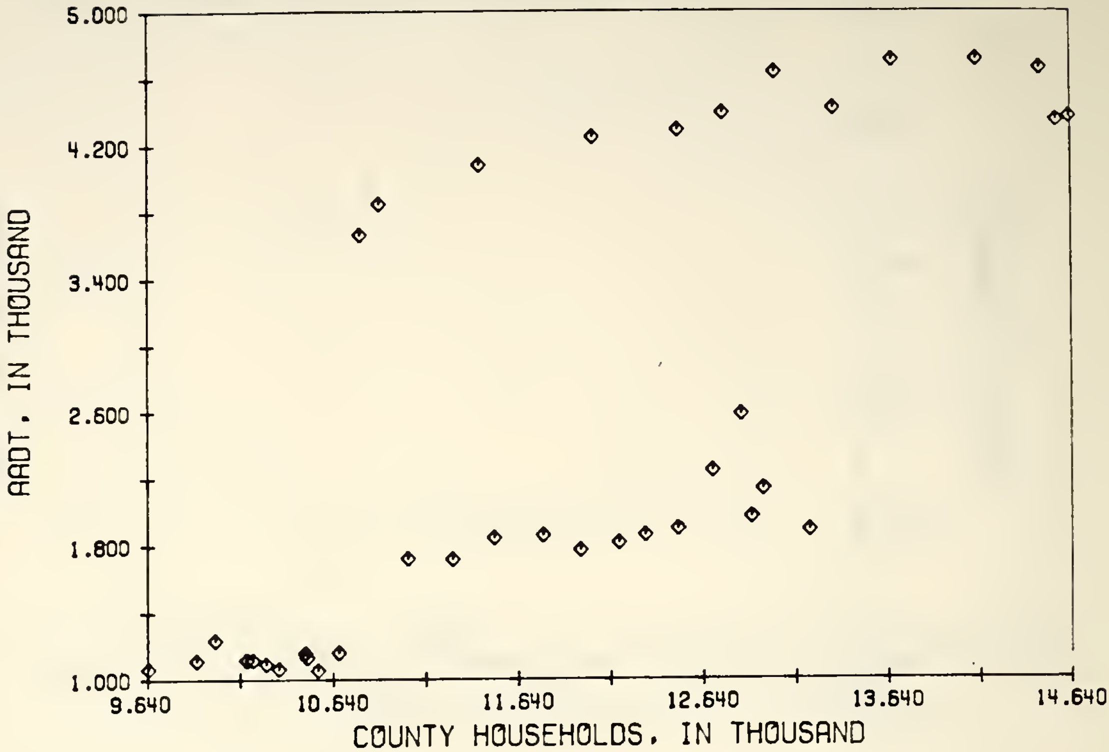
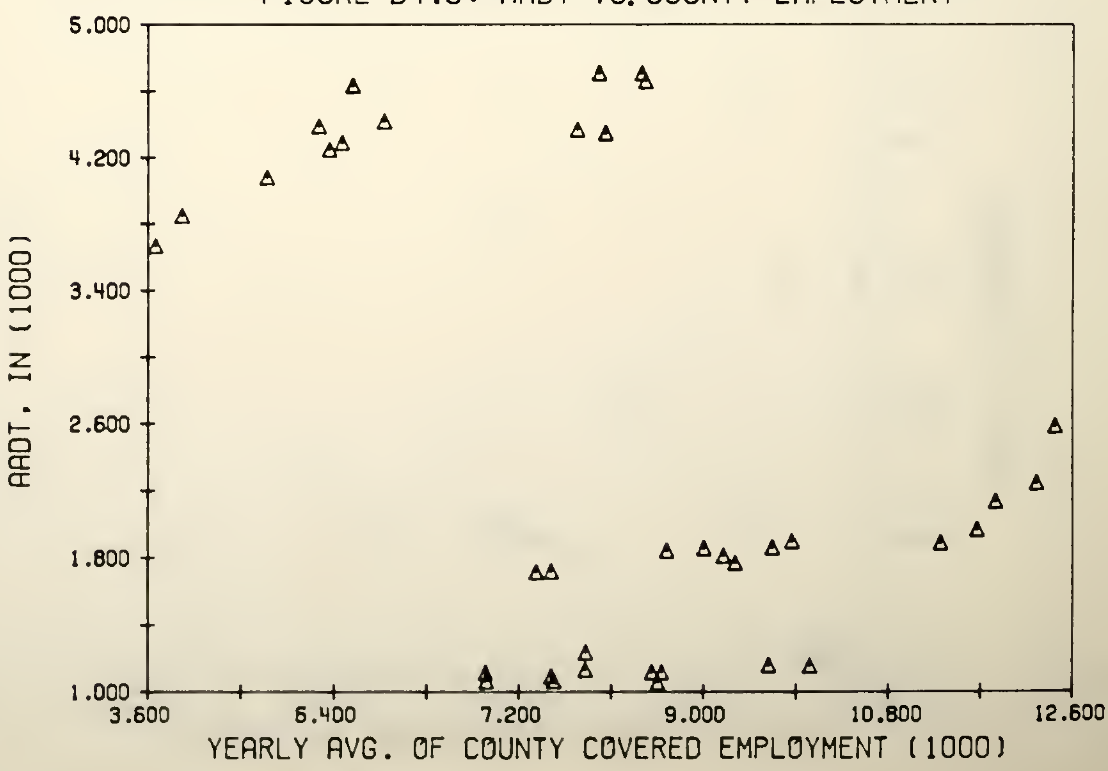


FIGURE B4.6: AADT VS. COUNTY EMPLOYMENT



Appendix C

Residual Plots:

Aggregate Analysis

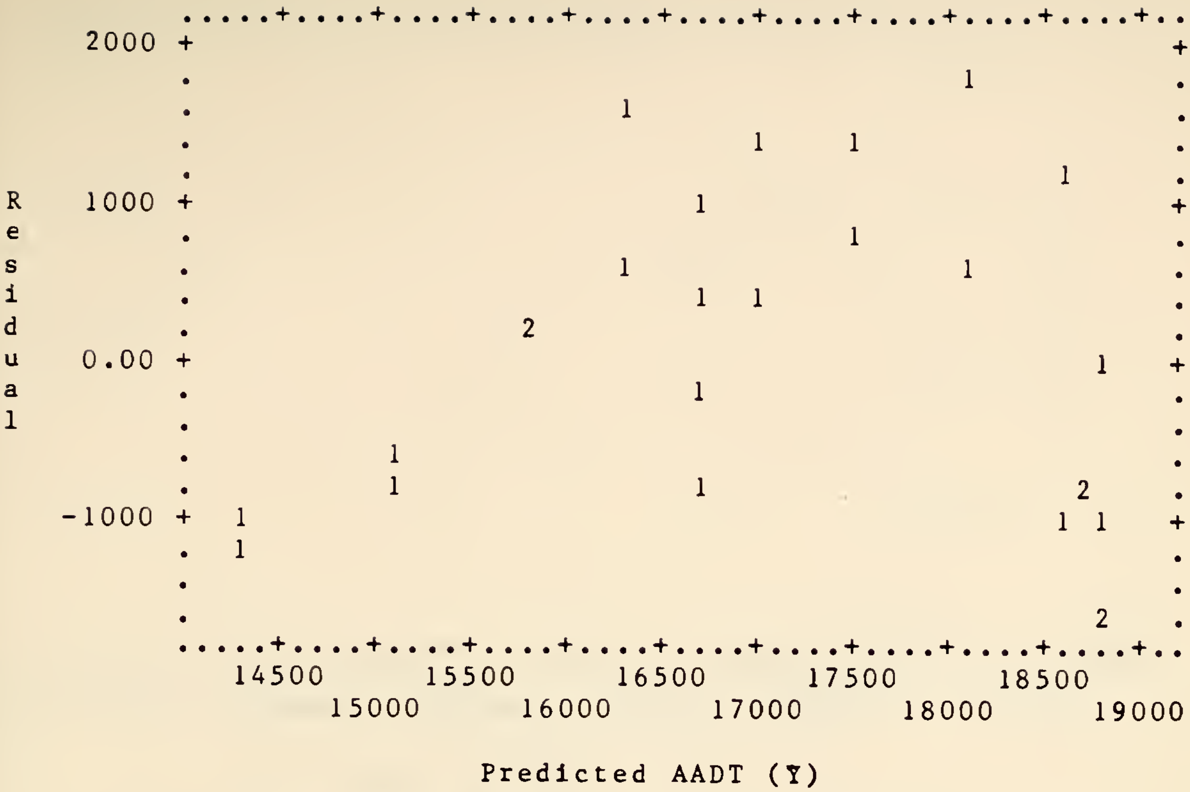


Figure C1.1: Residual Plot against Ŷ
(Rural Interstate)

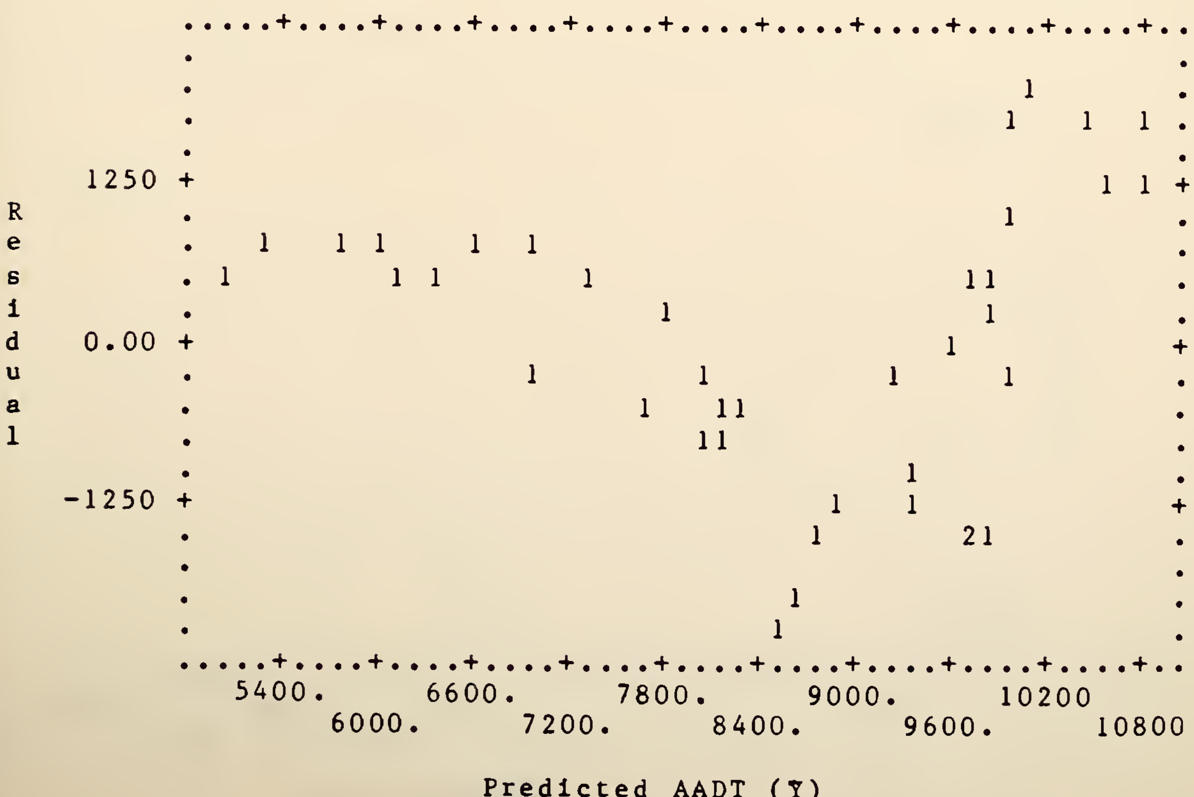


Figure C1.2: Residual Plot against Ŷ
(Rural Principal Arterial)

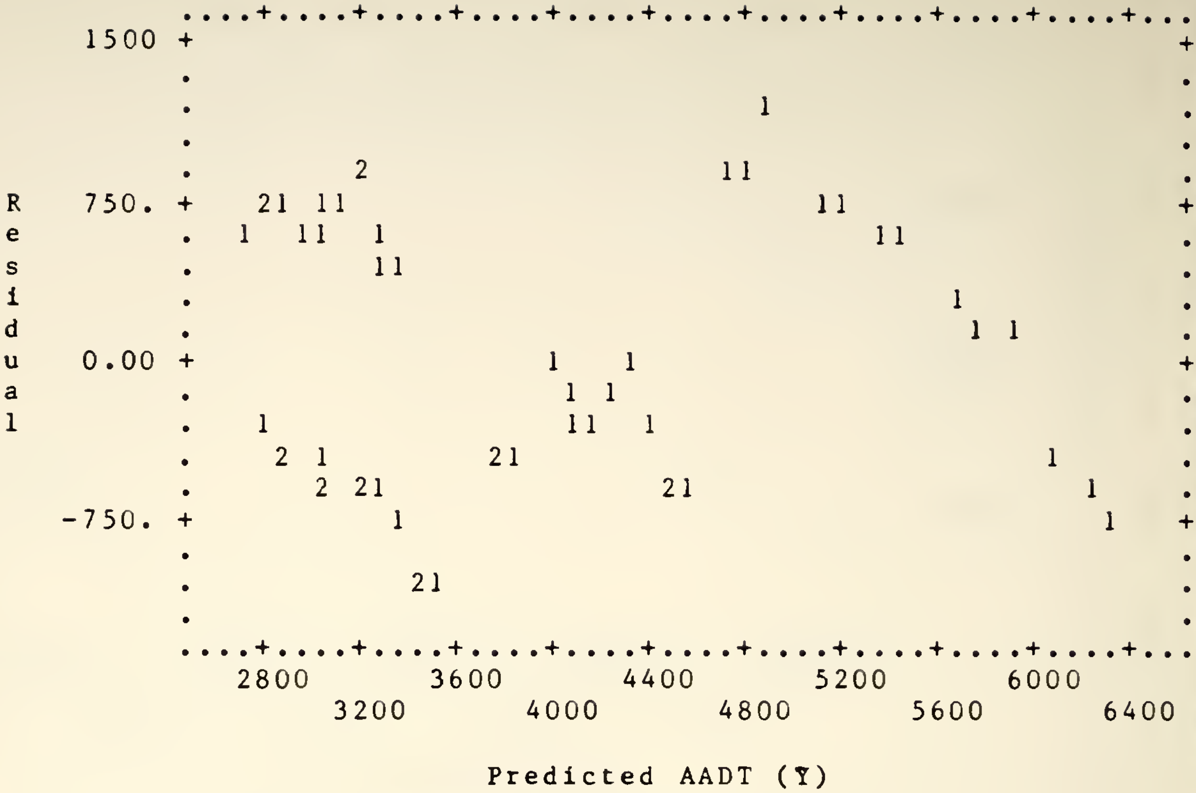


Figure C1.3: Residual Plot against Ŷ
(Rural Minor Arterial)

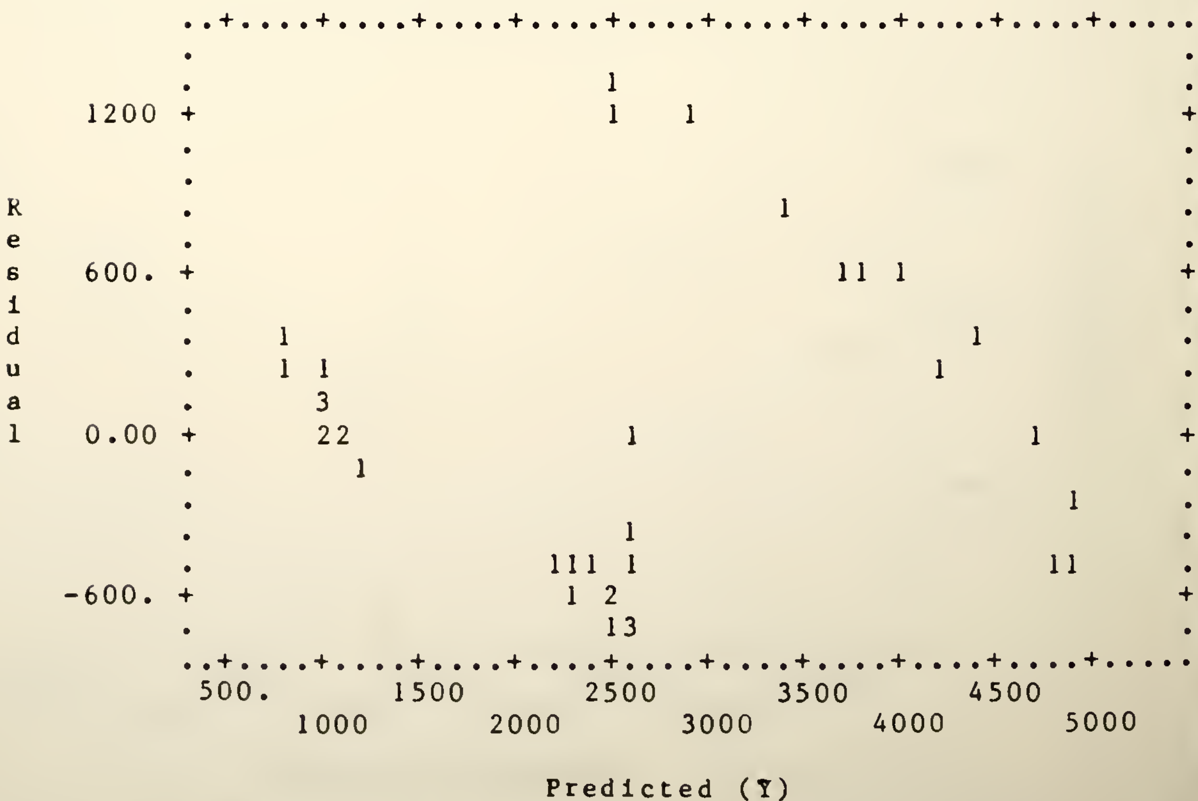


Figure C1.4: Residual Plot against Ŷ
(Rural Major Collector)

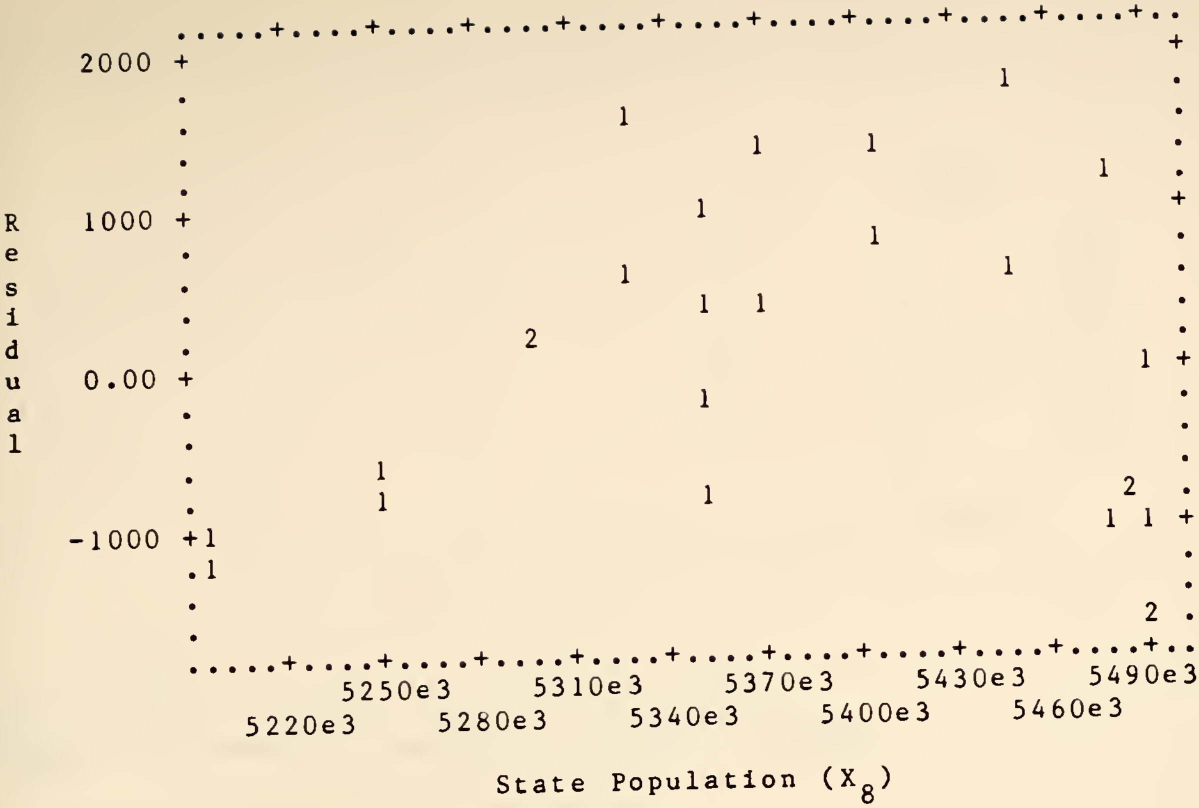


Figure C2.1: Residual Plot against X_8
(Rural Interstate)

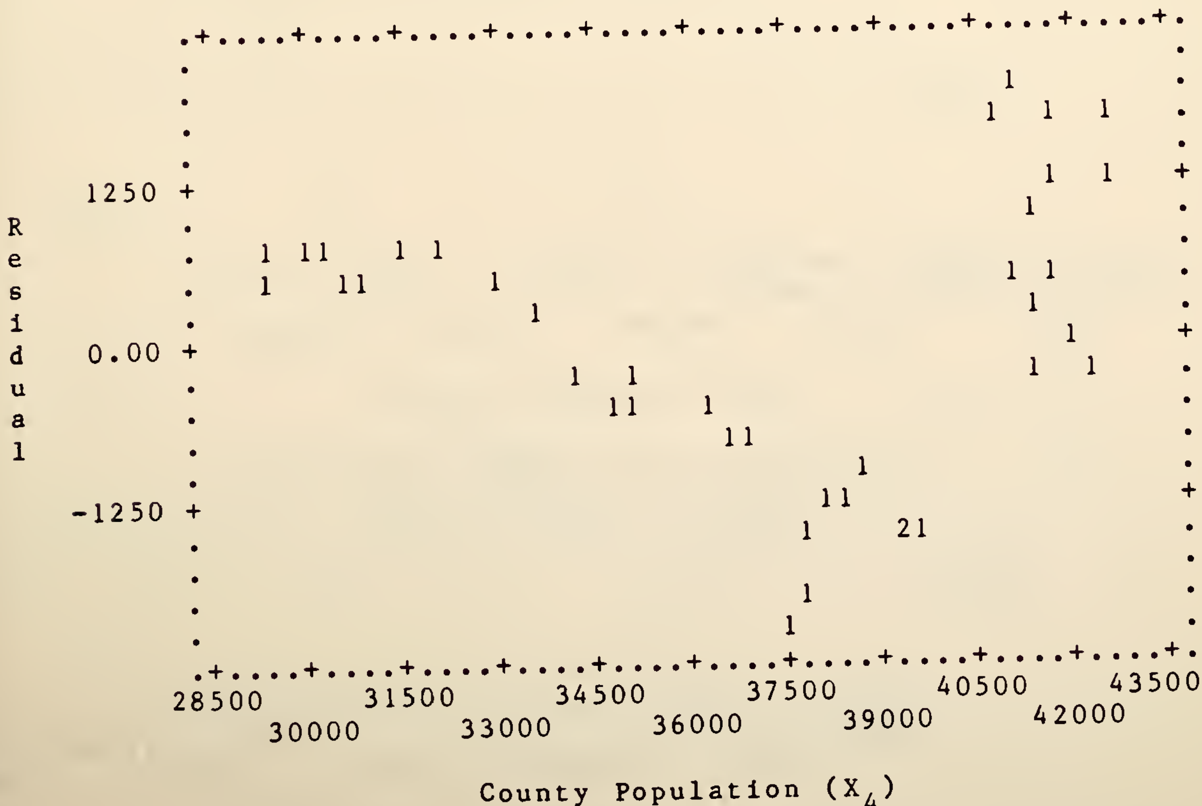


Figure C2.2.1: Residual Plot against X_4
(Rural Principal Arterial)

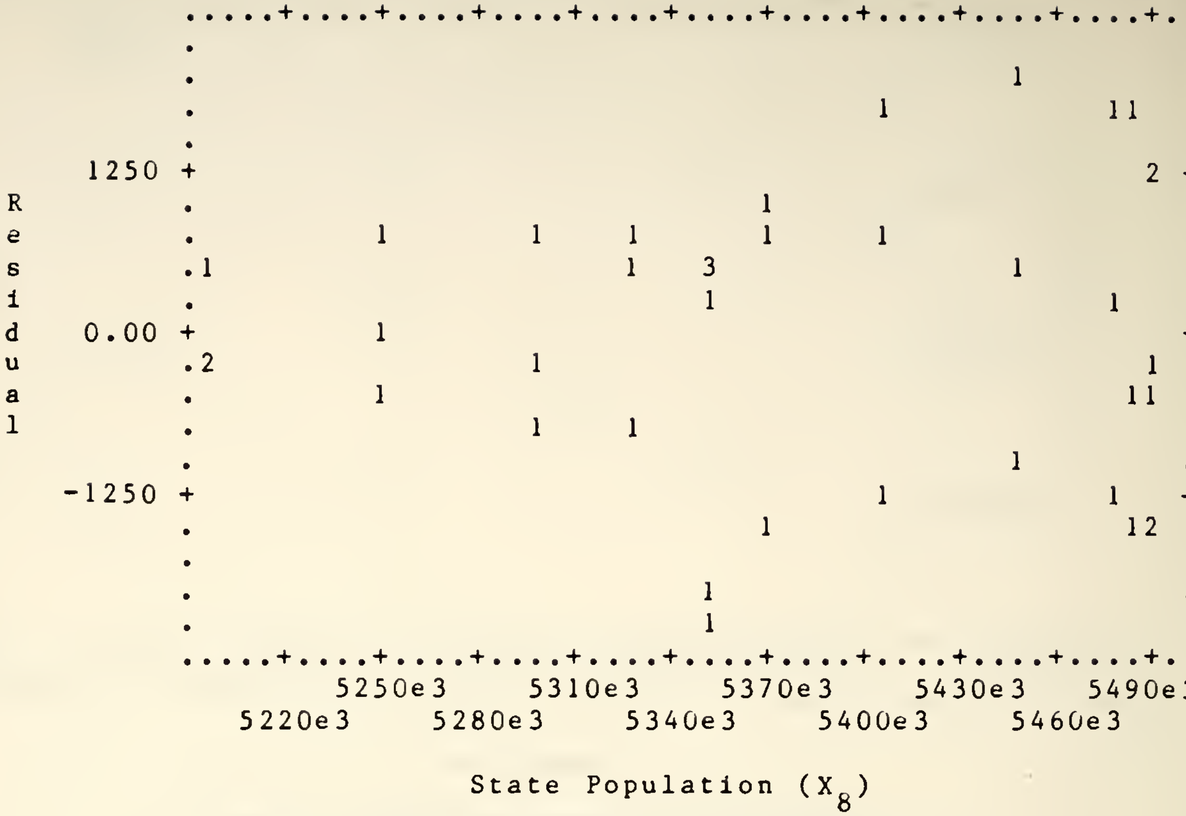


Figure C2.2.2: Residual Plot against X_8
(Rural Principal Arterial)

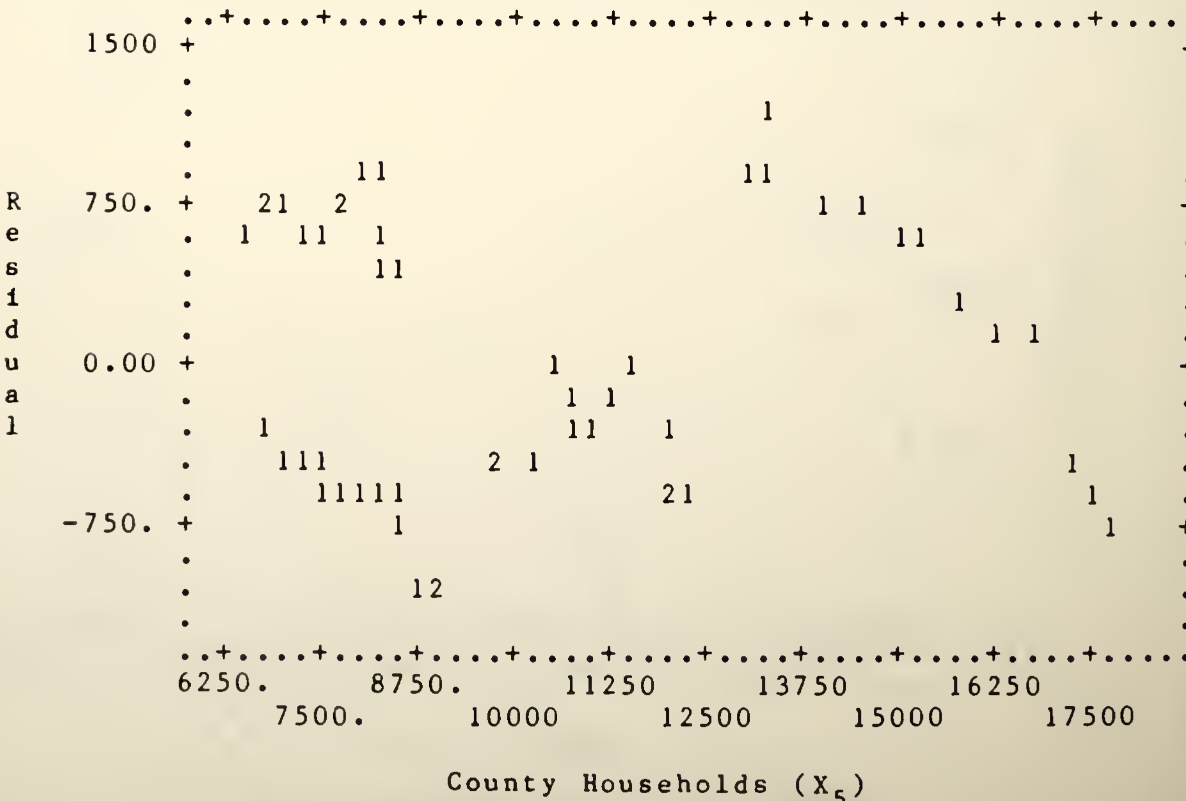


Figure C2.3: Residual Plot against X_5
(Rural Minor Arterial)

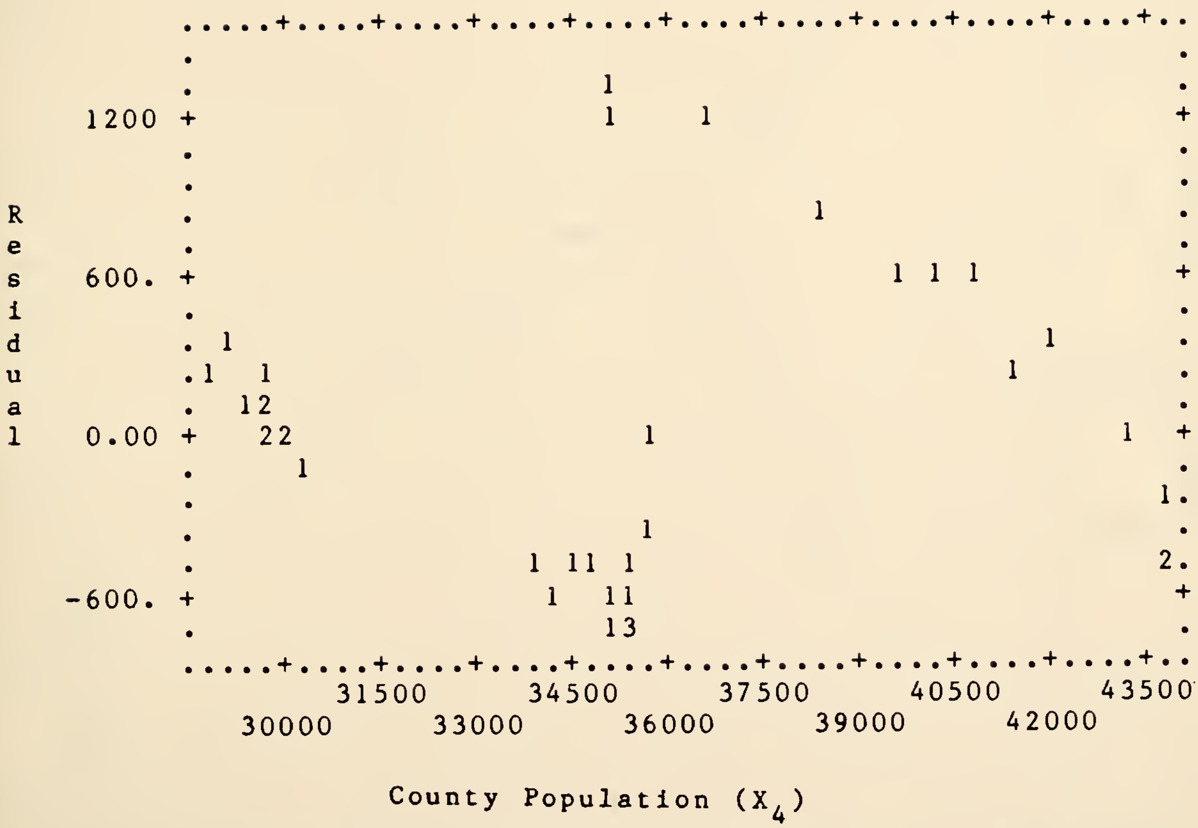


Figure C2.4: Residual Plot against X_4
(Rural Major Collector)

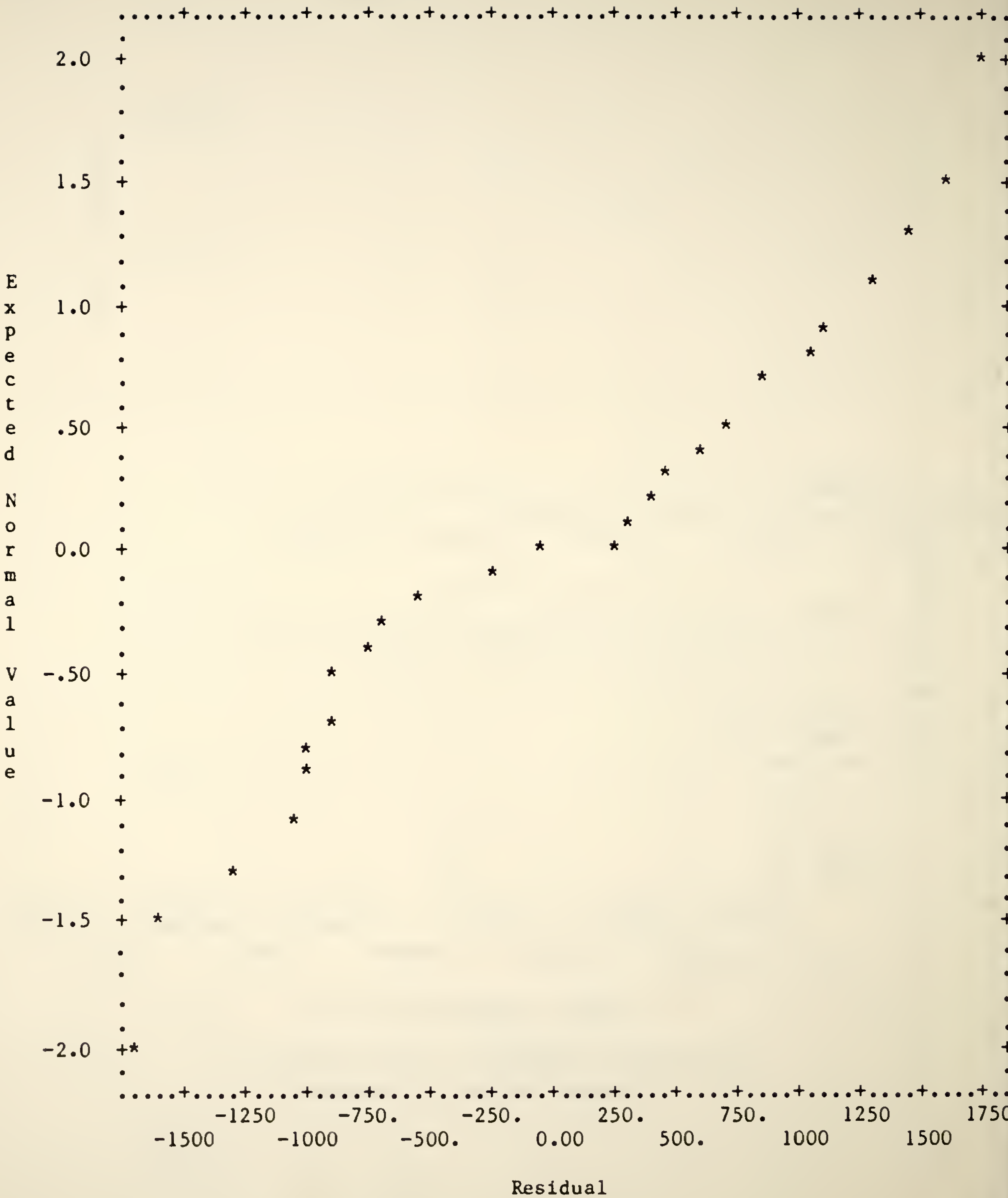


Figure C3.1: Normal Probability Plot of Residuals
(Rural Interstate)

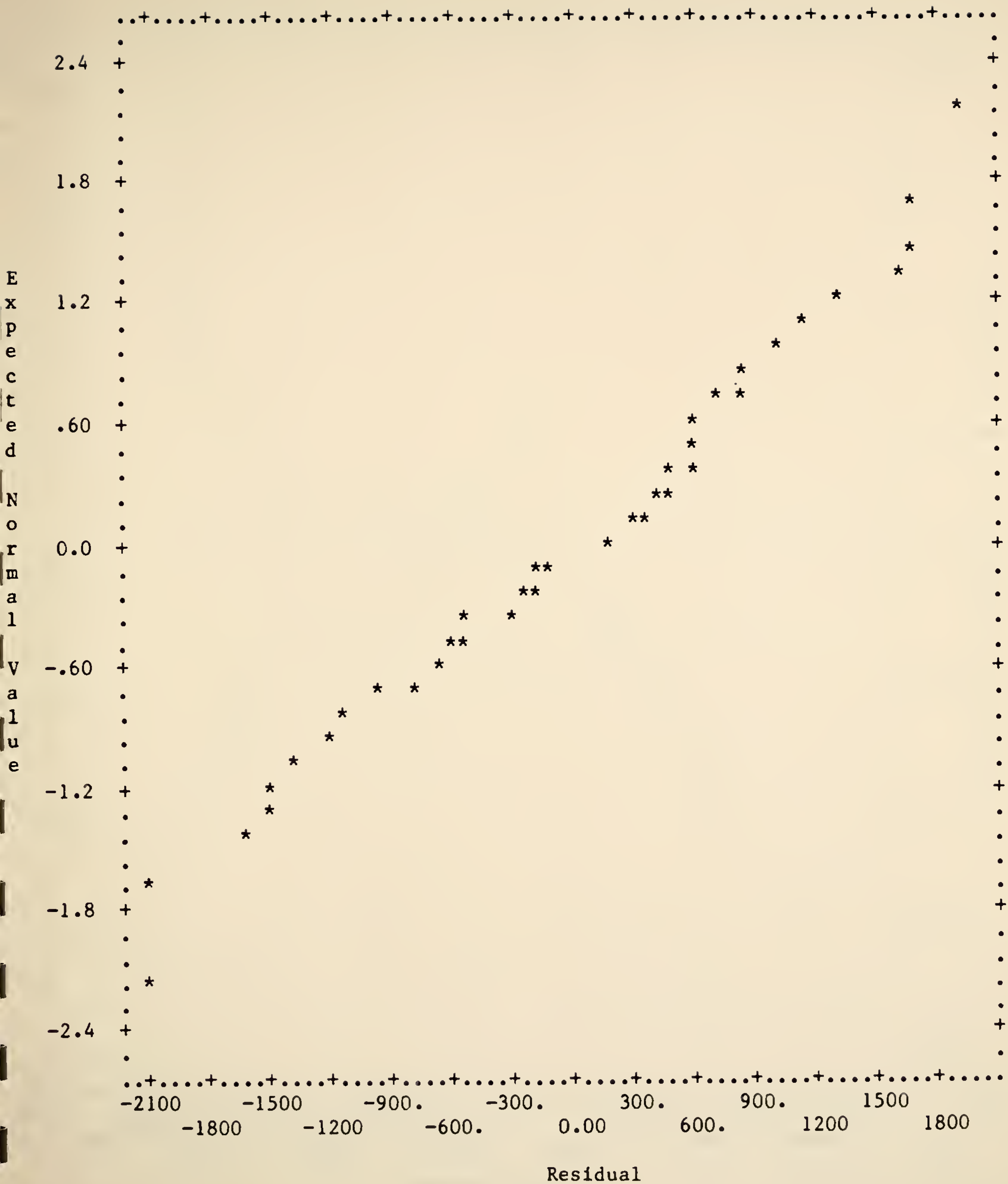


Figure C3.2: Normal Probability Plot of Residuals
(Rural Principal Arterial)

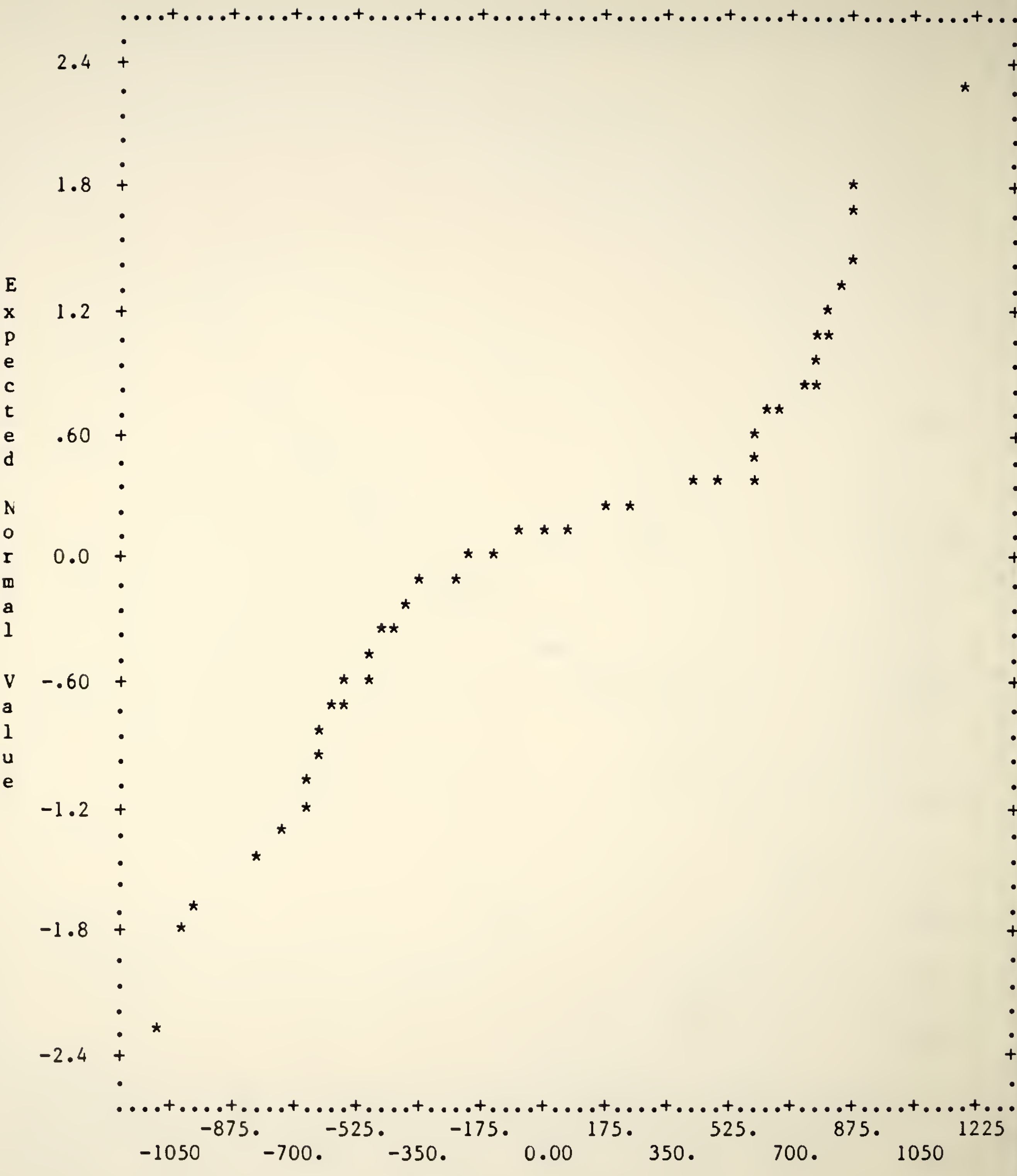


Figure C3.3: Normal Probability Plot of Residuals
(Rural Minor Arterial)

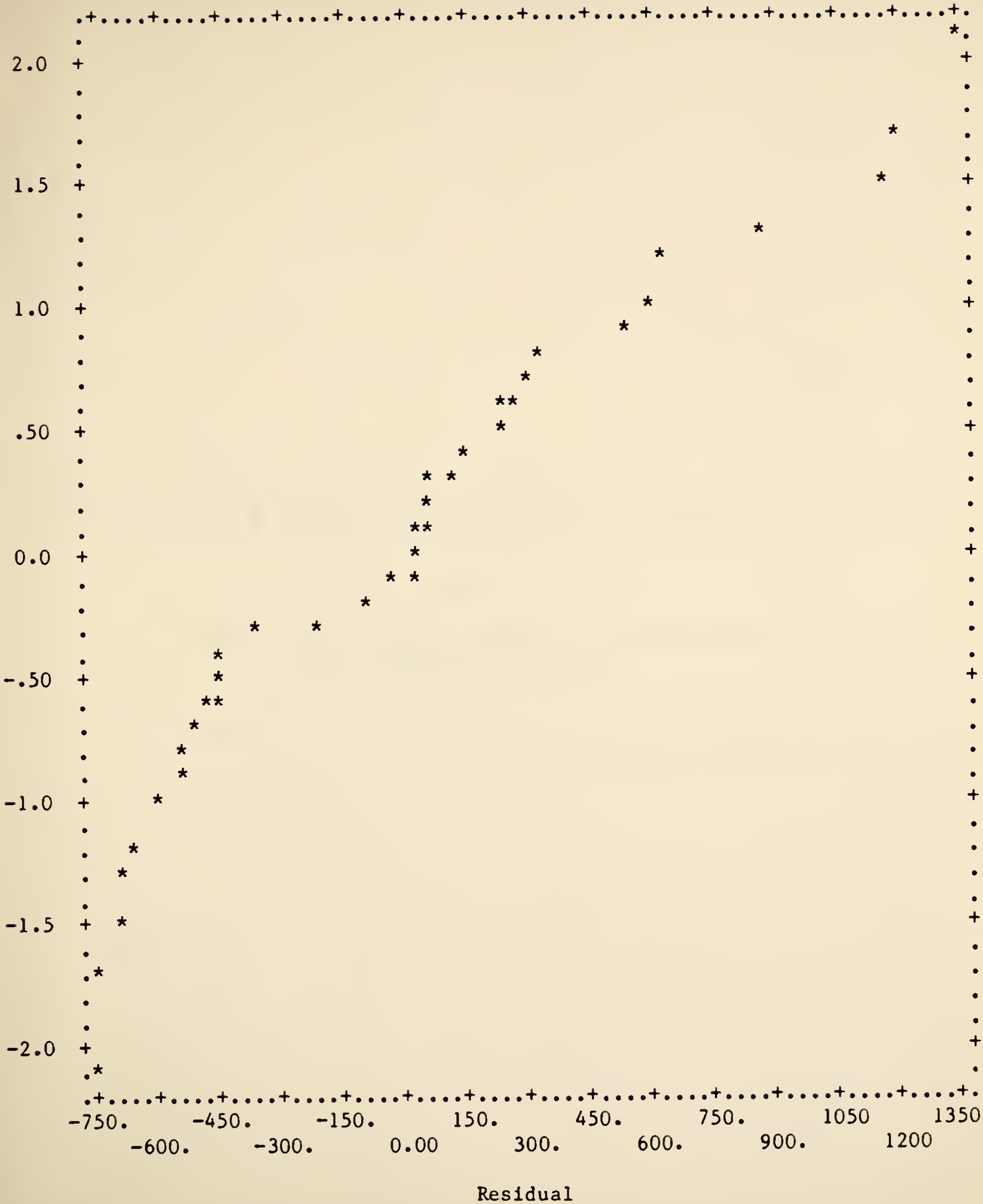


Figure C3.4: Normal Probability Plot of Residuals
(Rural Major Collector)

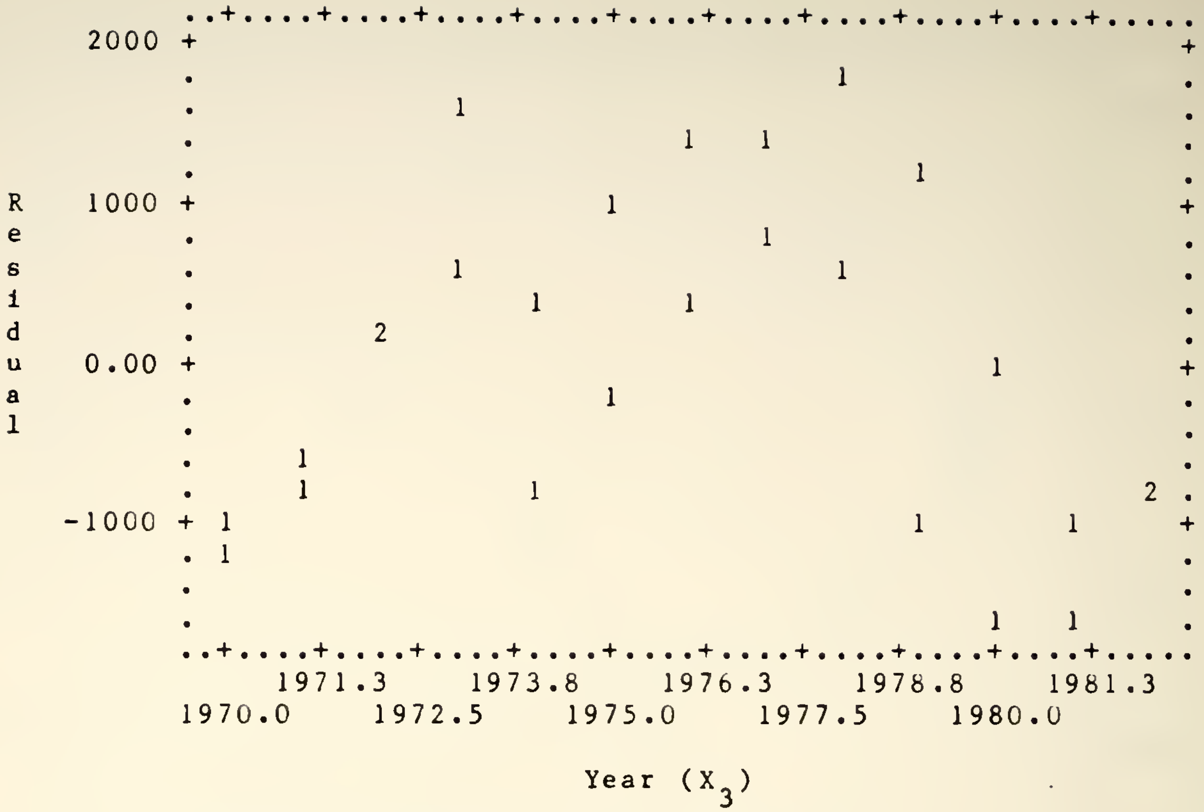


Figure C4.1: Residual Plot against X_3
(Rural Interstate)

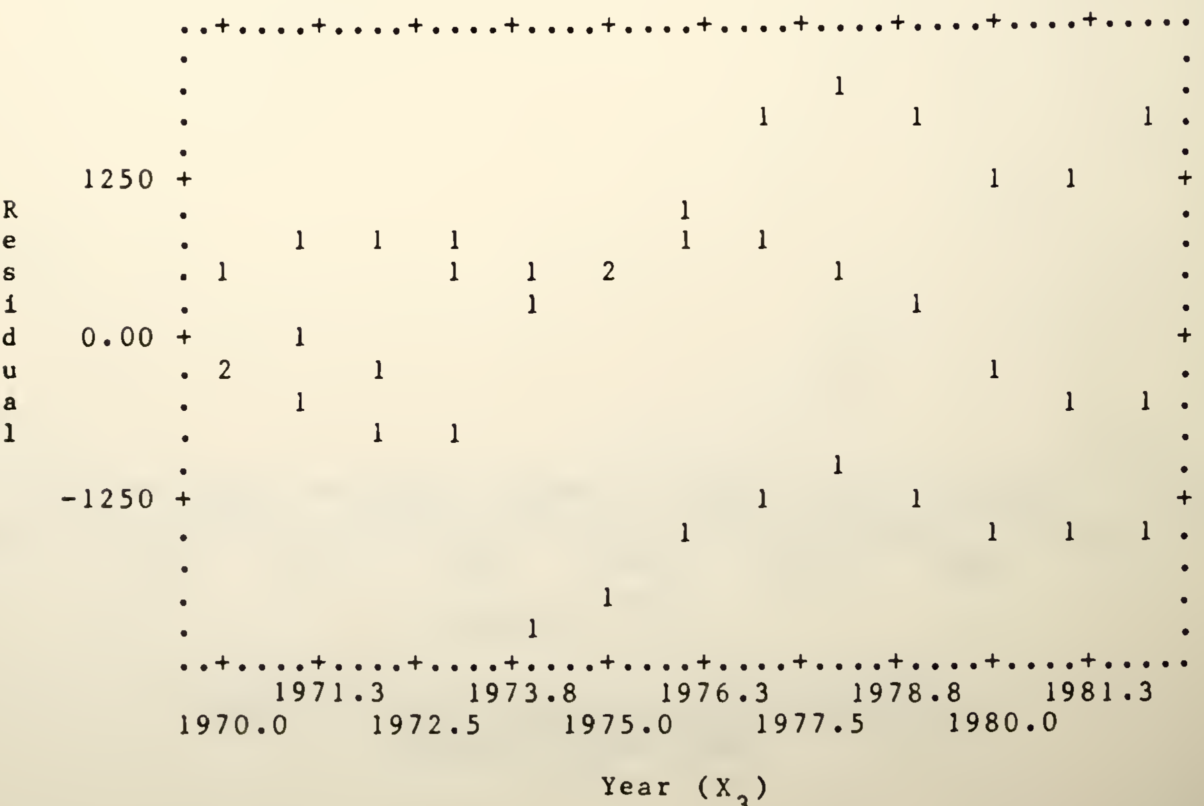


Figure C4.2: Residual Plot against X_3
(Rural Principal Arterial)

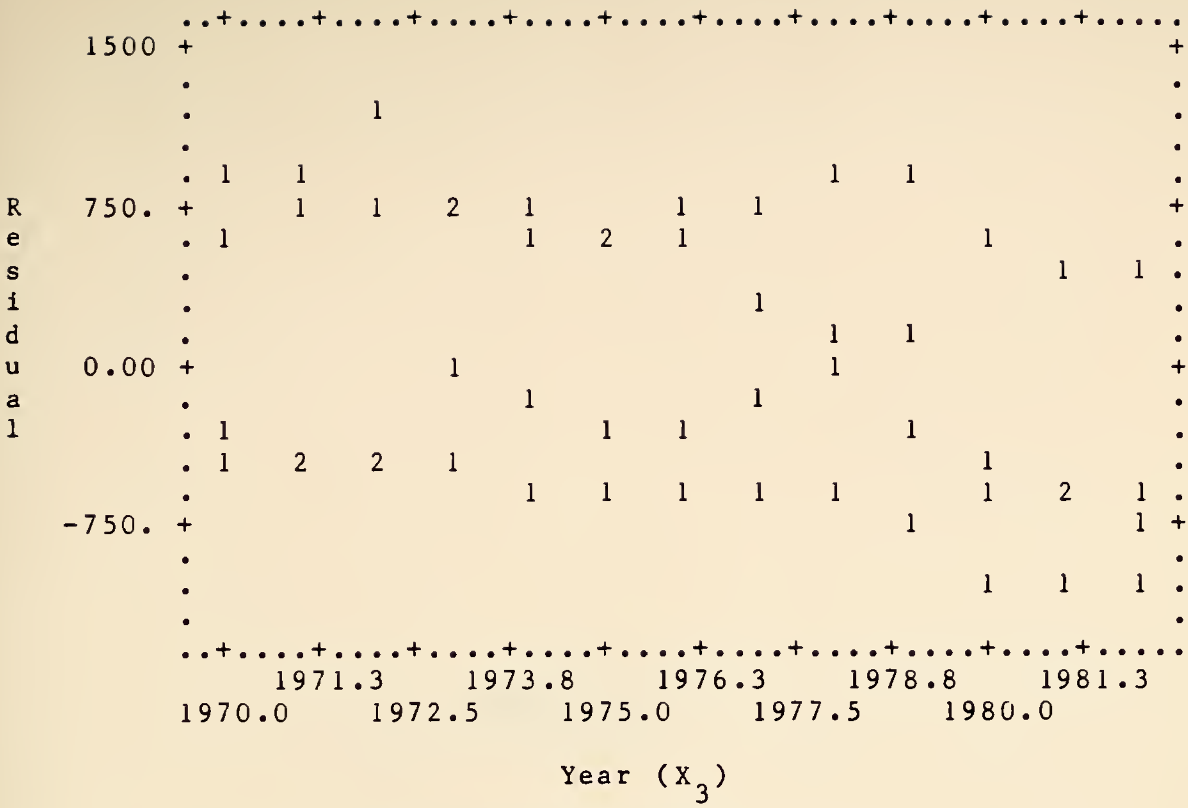


Figure C4.3: Residual Plot against X_3
(Rural Minor Arterial)

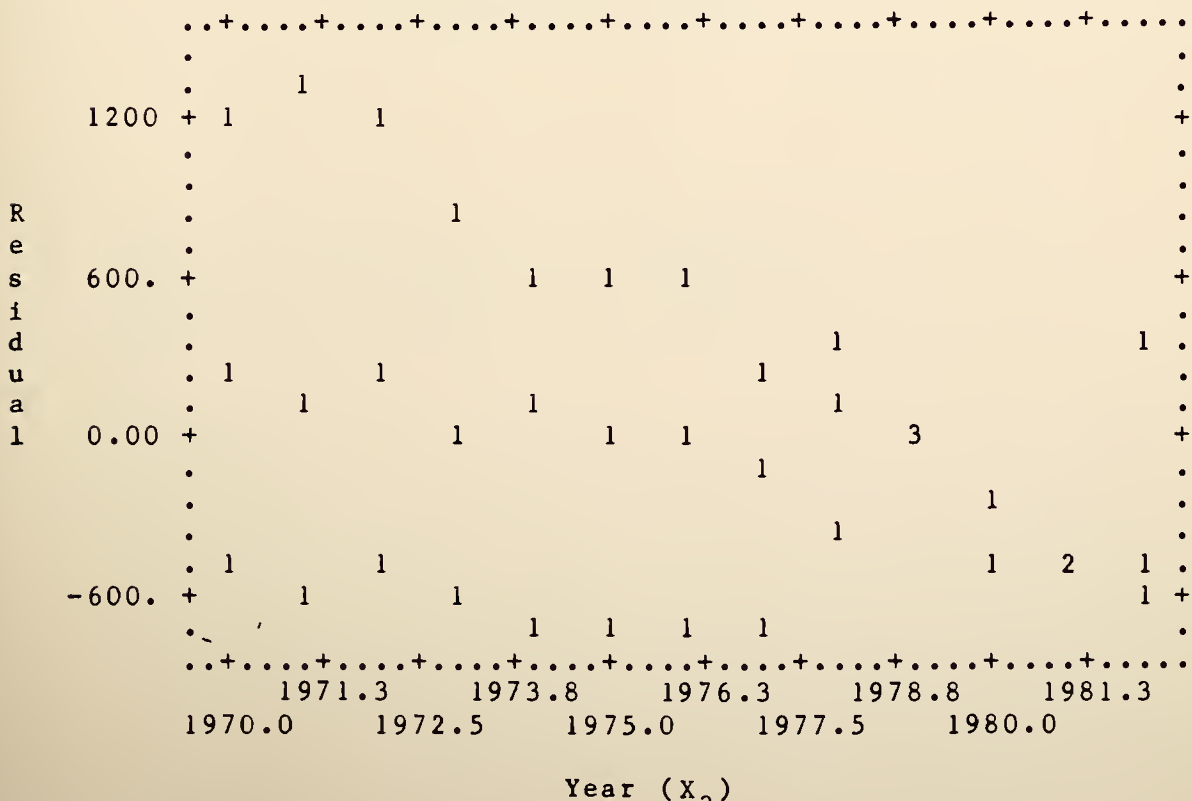


Figure C4.4: Residual Plot against X_3
(Rural Major Collector)

Appendix D

Scatter Plots:

Disaggregate Analysis

1. Station 68A: Figure D1 to Figure D11
2. Station 7047A: Figure D12 to Figure D17

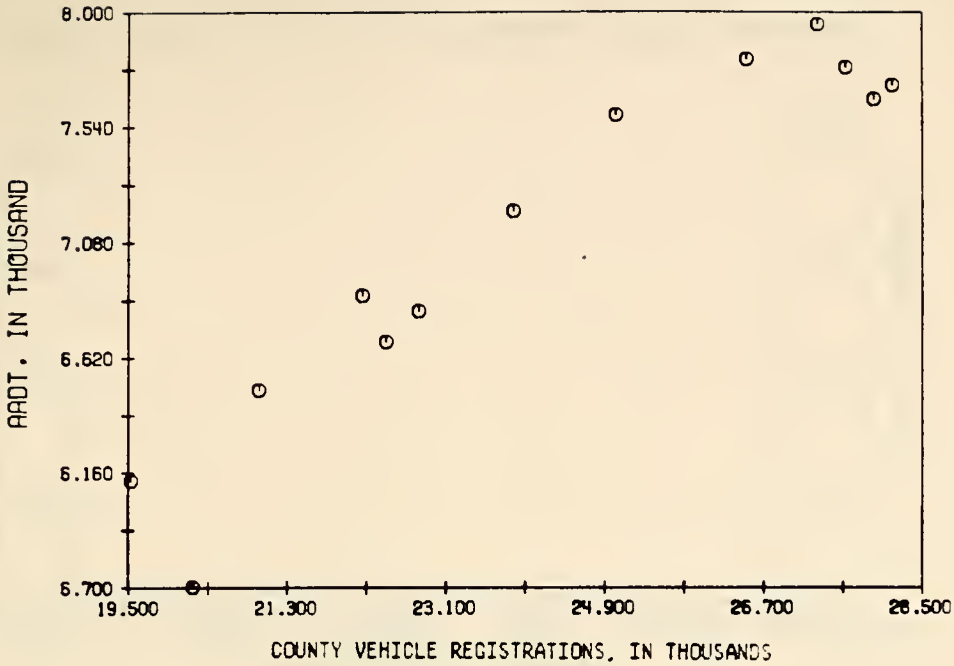


Figure D1: AADT vs. County Vehicle Registrations
(Station 68A)

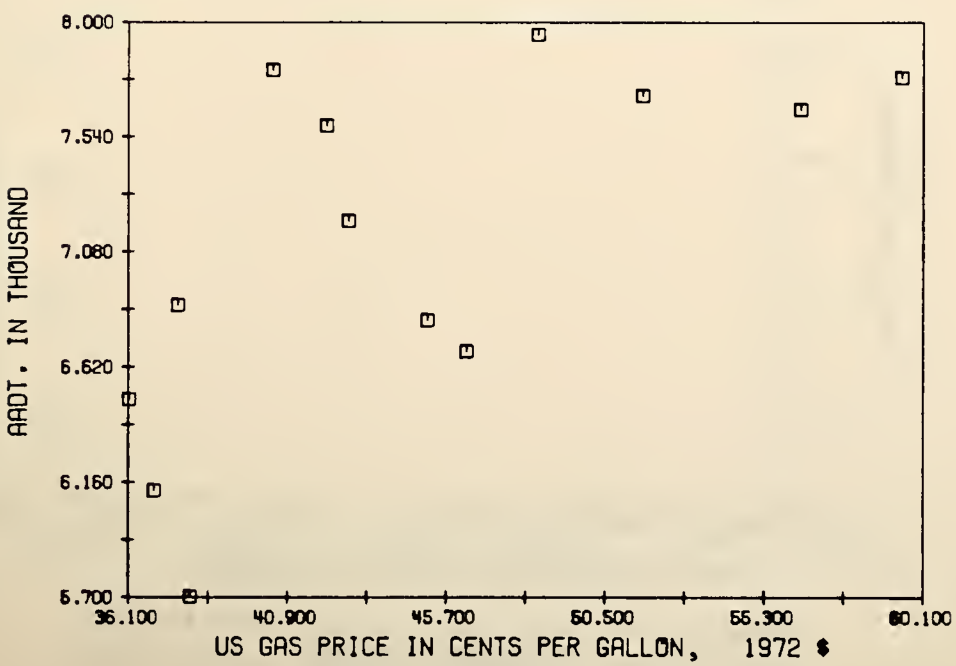


Figure D2: AADT vs. US Gasoline Price
(Station 68A)

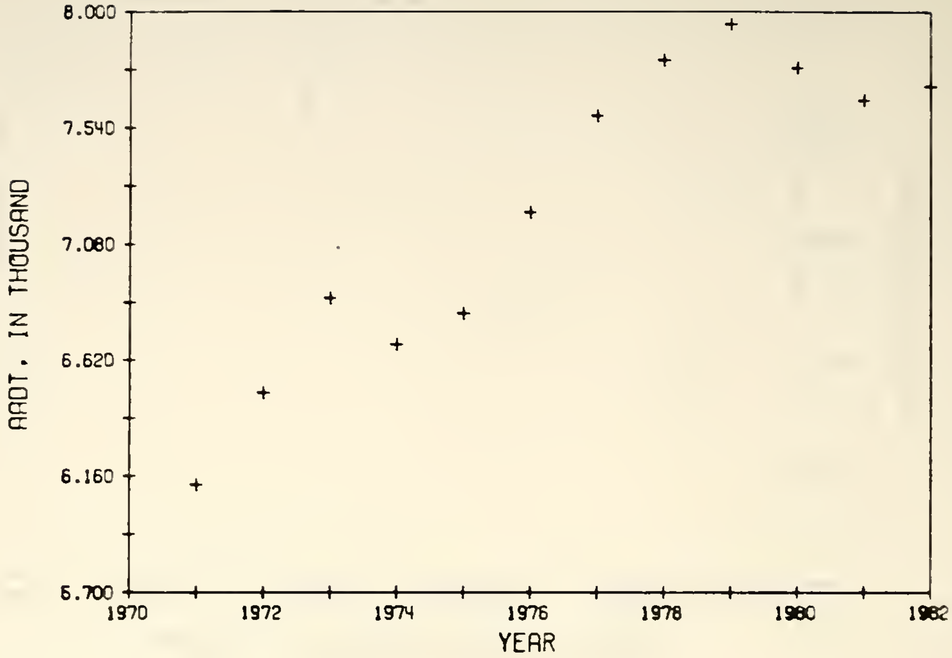


Figure D3: AADT vs. Year
(Station 68A)

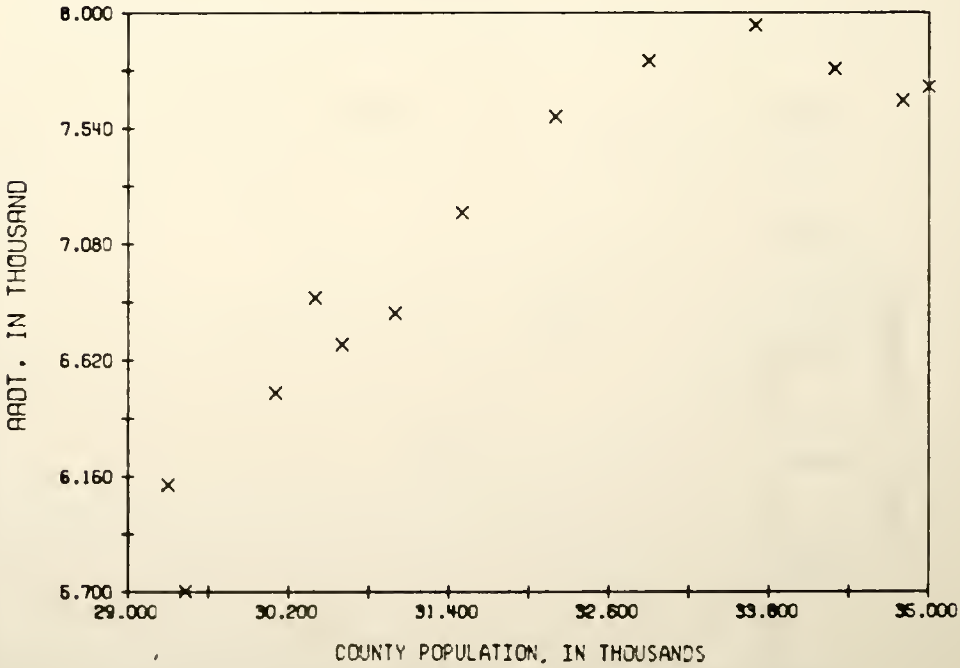


Figure D4: AADT vs. County Population
(Station 68A)

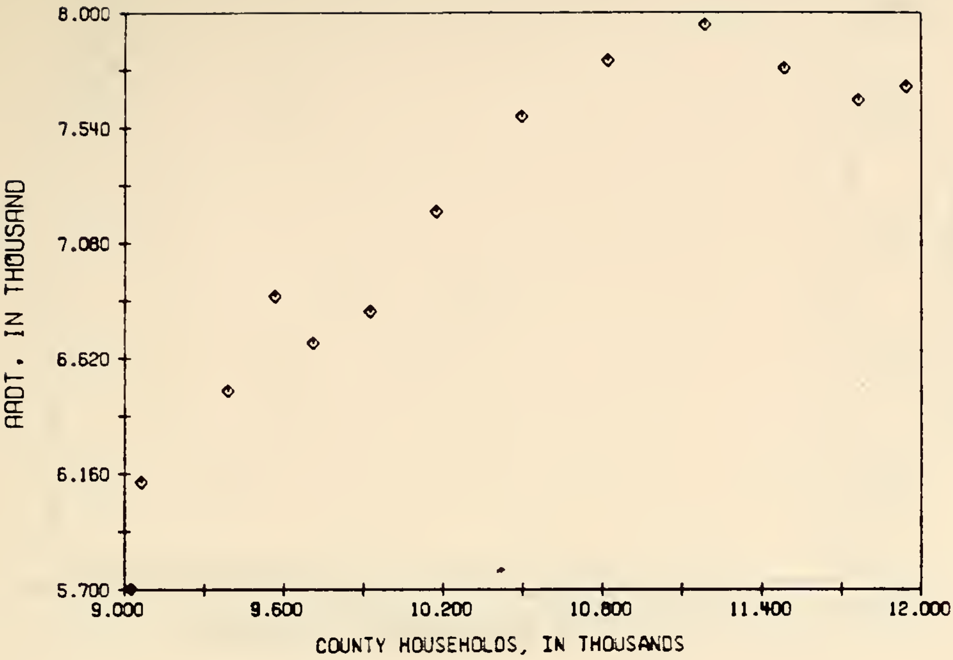


Figure D5: AADT vs. County Households
(Station 68A)

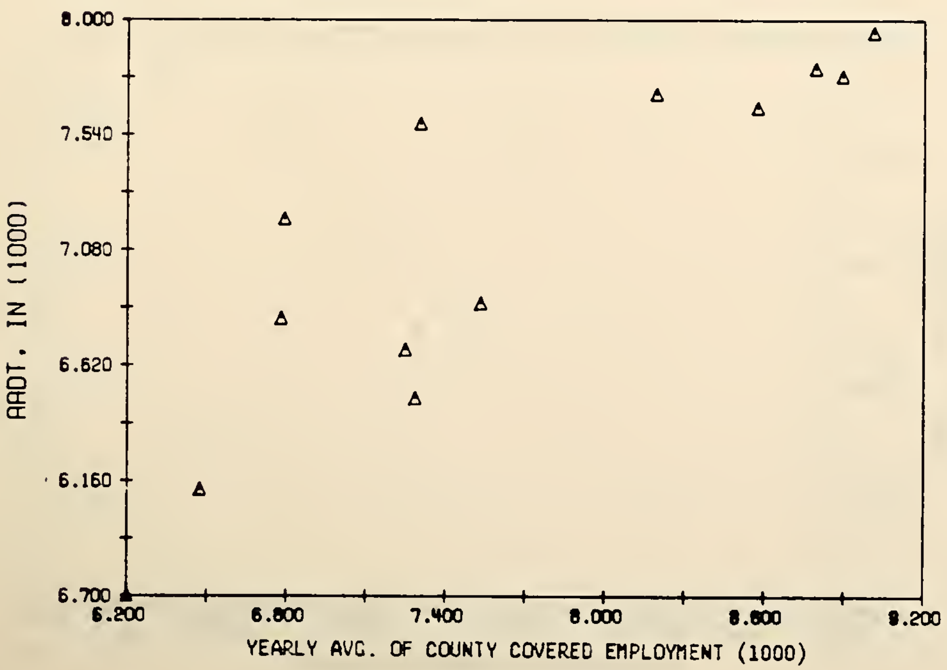


Figure D6: AADT vs. County Employment
(Station 68A)

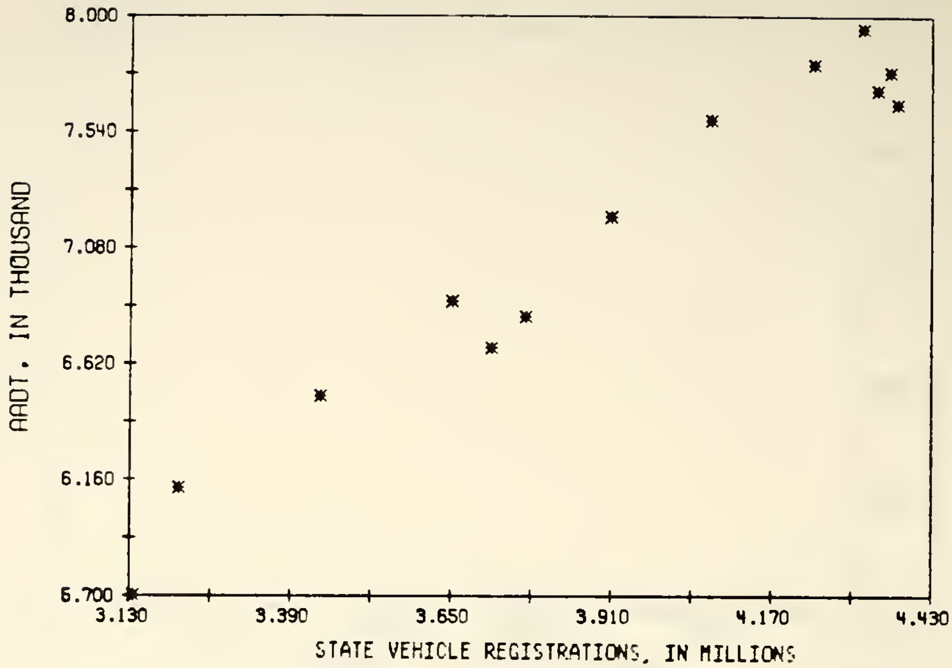


Figure D7: AADT vs. State Vehicle Registrations (Station 68A)

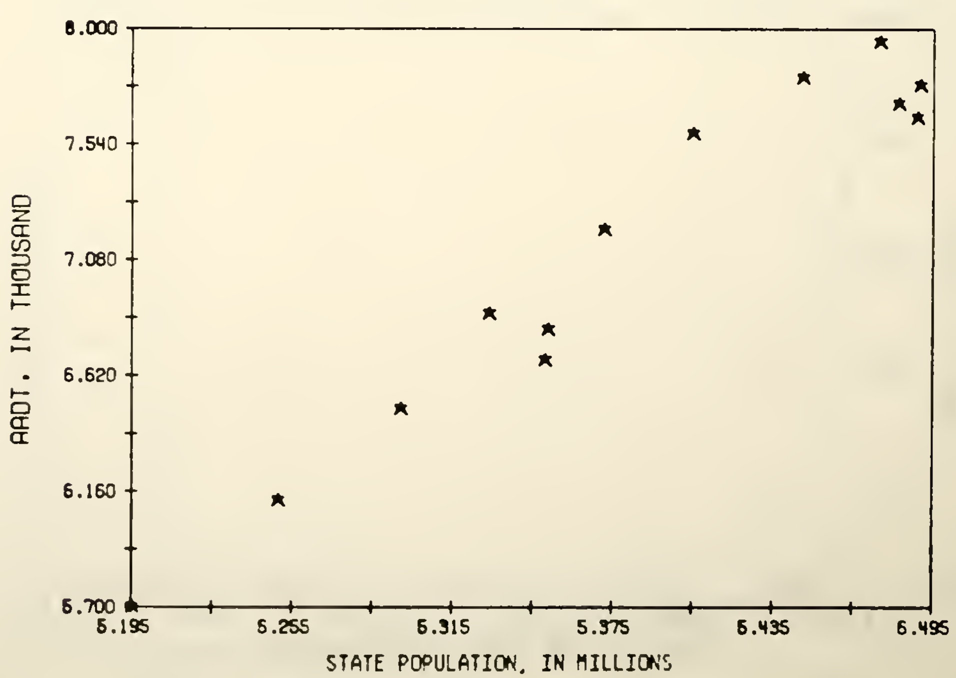


Figure D8: AADT vs. State Population (Station 68A)

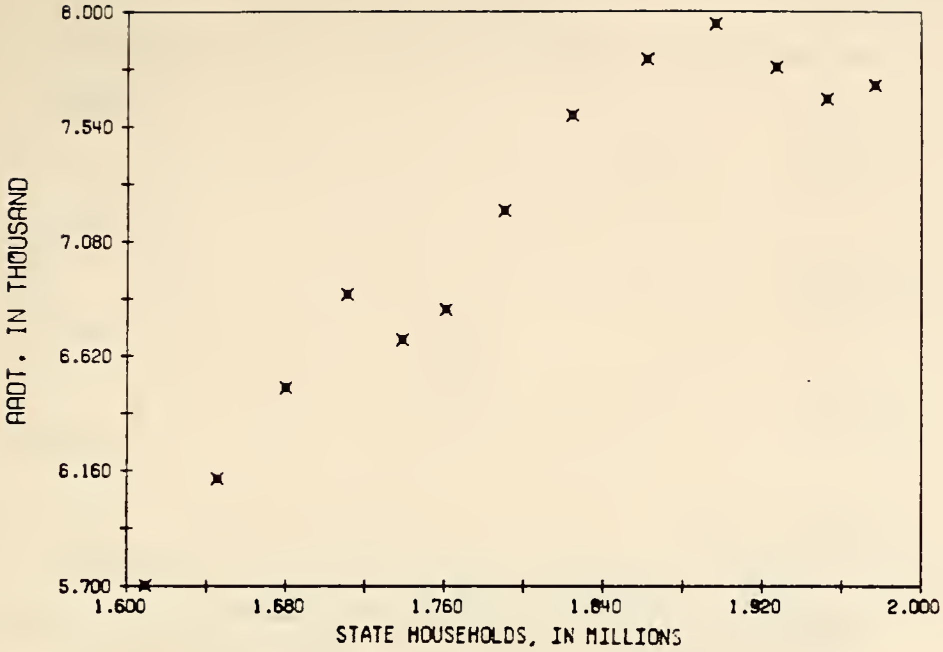


Figure D9: AADT vs. State Households
(Station 68A)

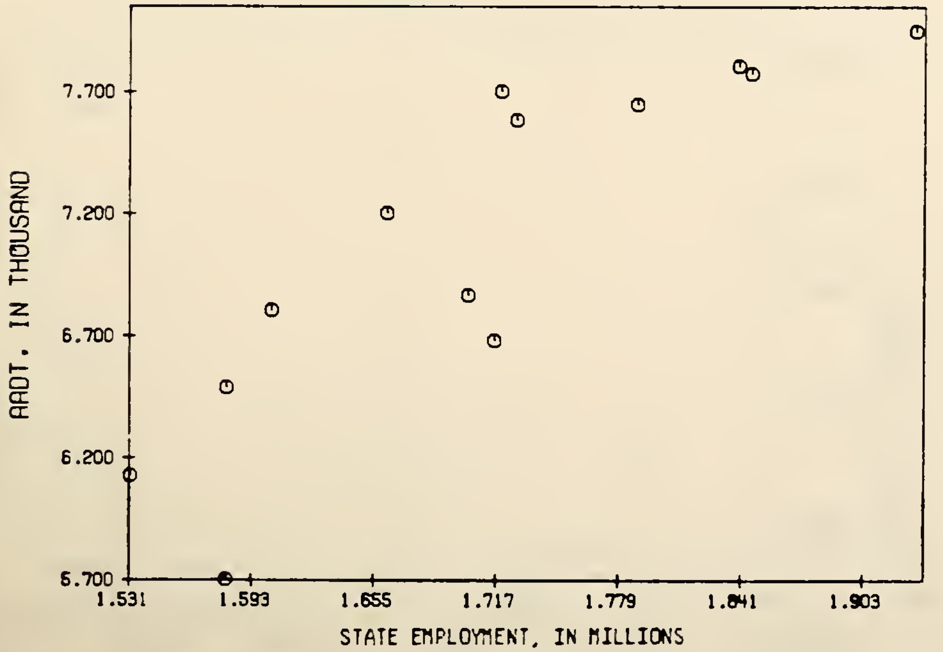


Figure D10: AADT vs. State Employment
(Station 68A)

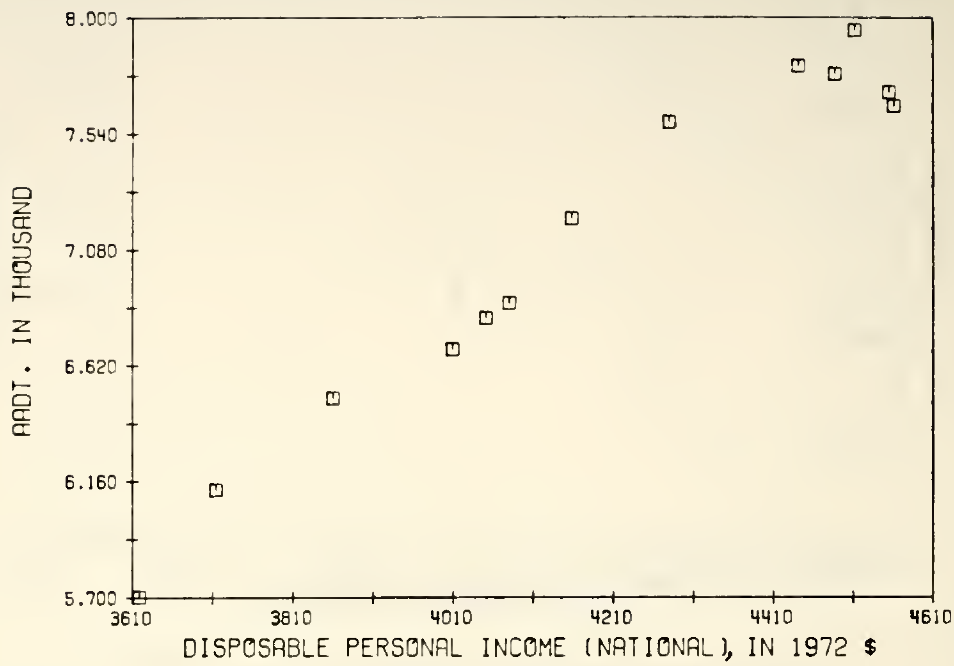


Figure D11: AADT vs. Per Capita National Income (Station 68A)

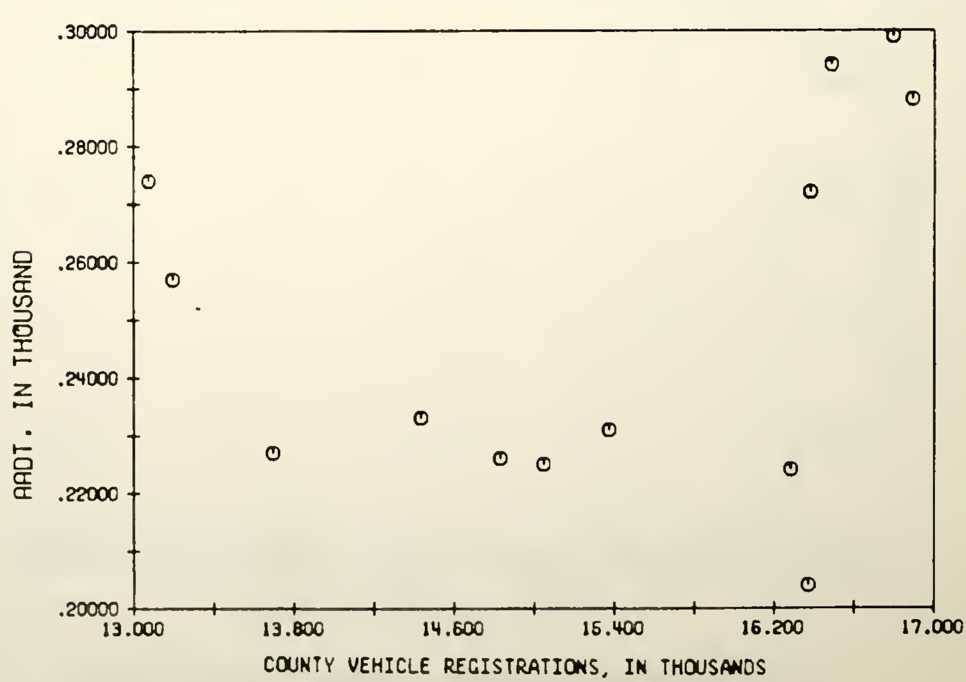


Figure D12: AADT vs. County Vehicle Registrations (Station 7047A)

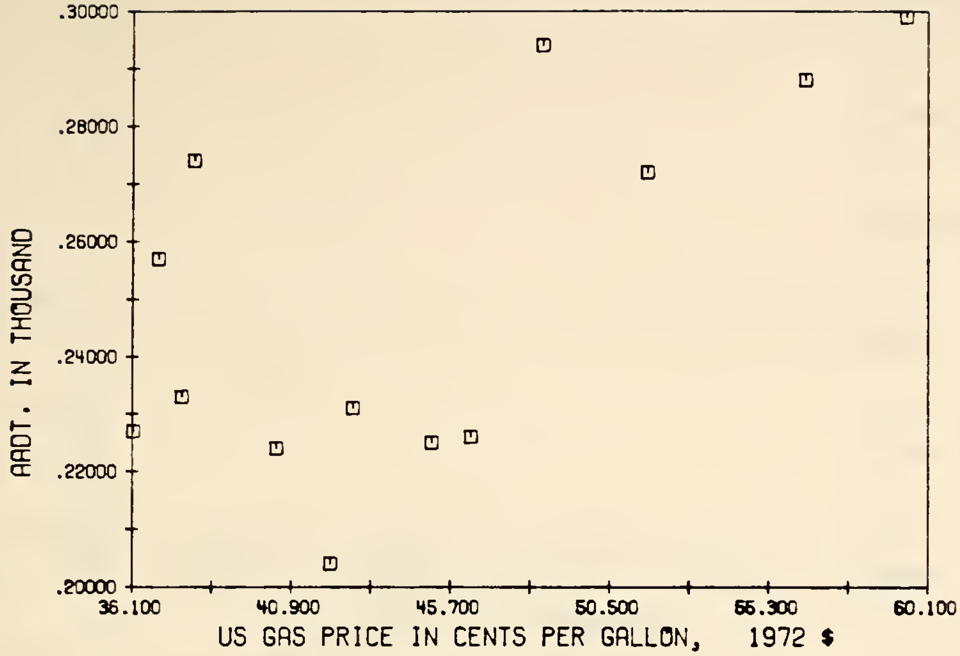


Figure D13: AADT vs. US Gasoline Price
(Station 7047A)

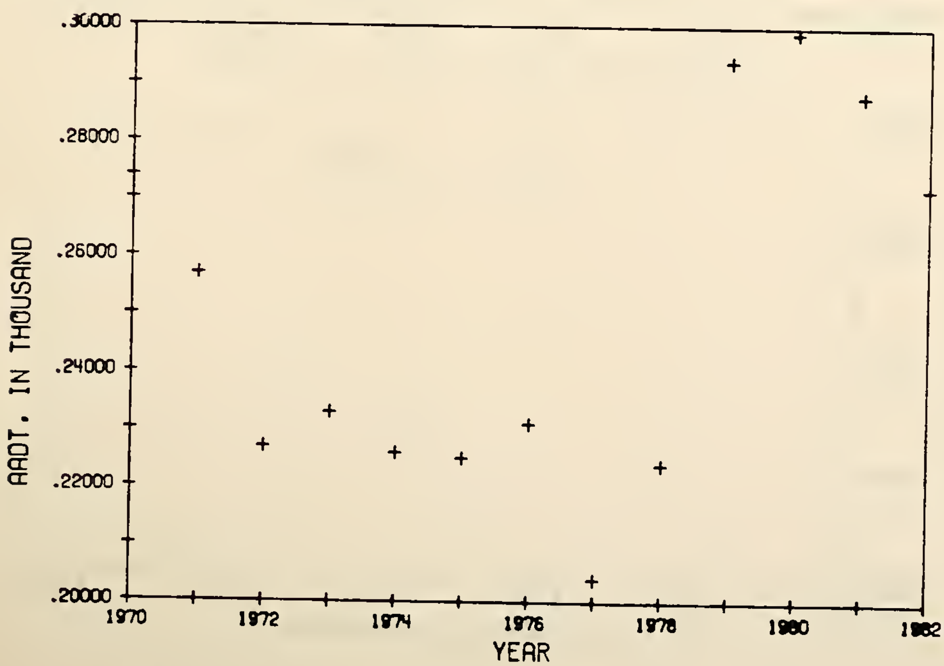


Figure D14: AADT vs. Year
(Station 7047A)

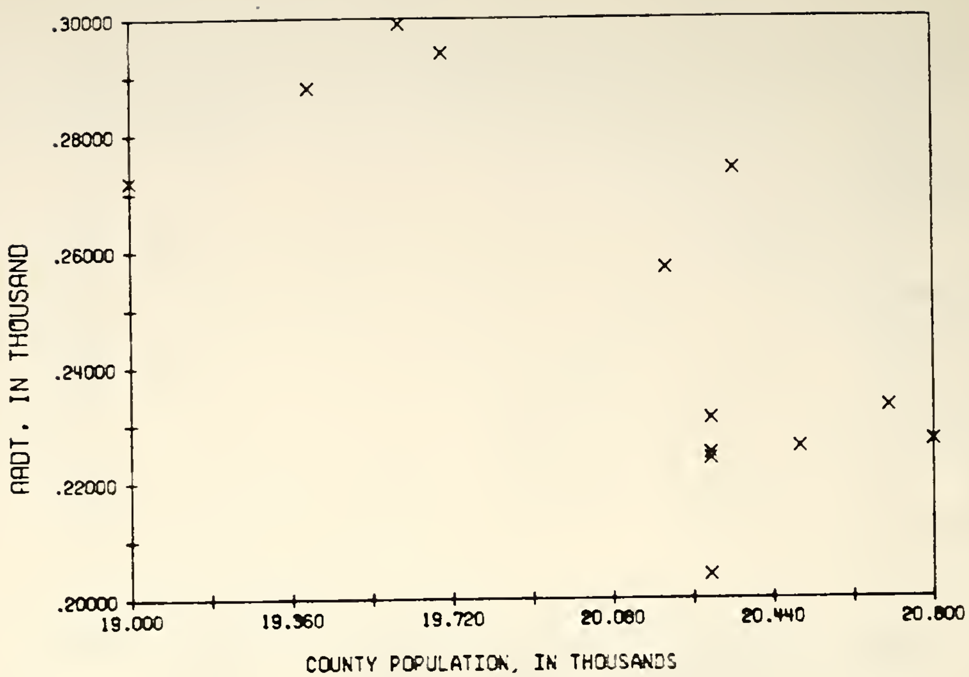


Figure D15: AADT vs. County Population
(Station 7047A)

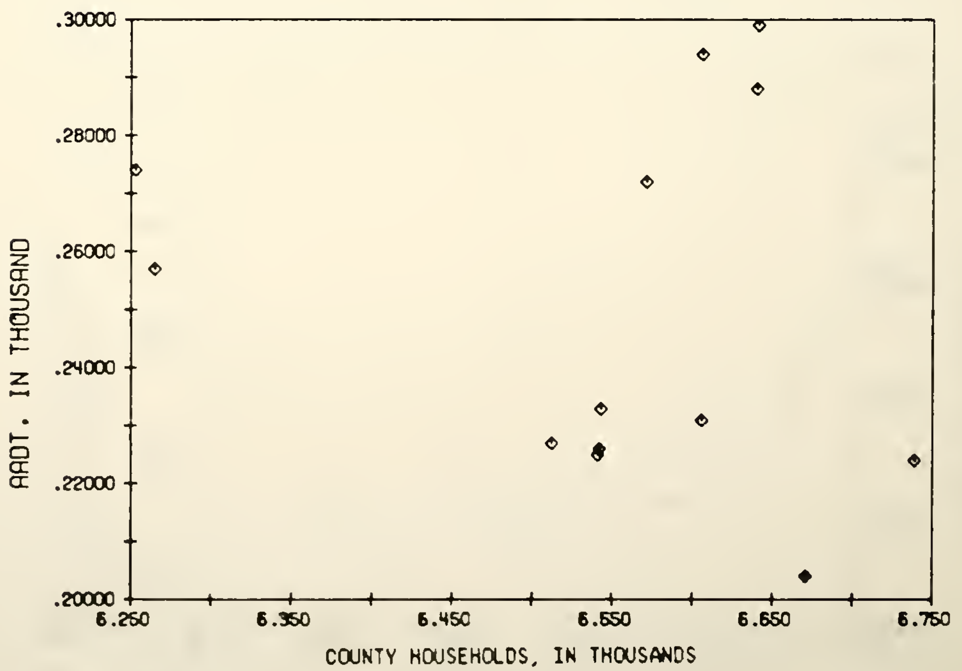


Figure D16: AADT vs. County Households
(Station 7047A)

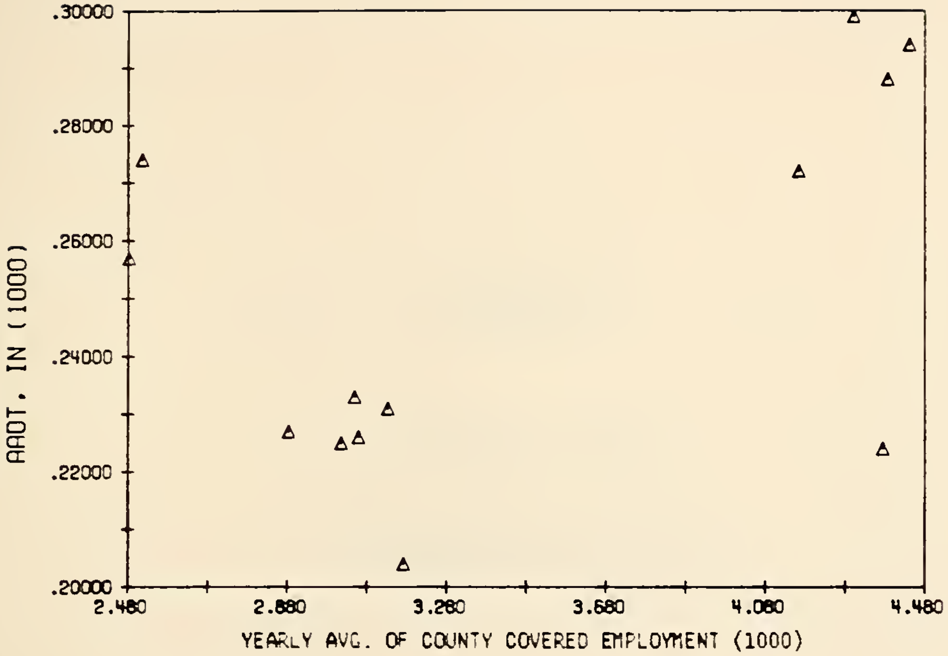


Figure D17: AADT vs. County Employment
(Station 7047A)

Appendix E

Residual Plots:

Disaggregate Analysis

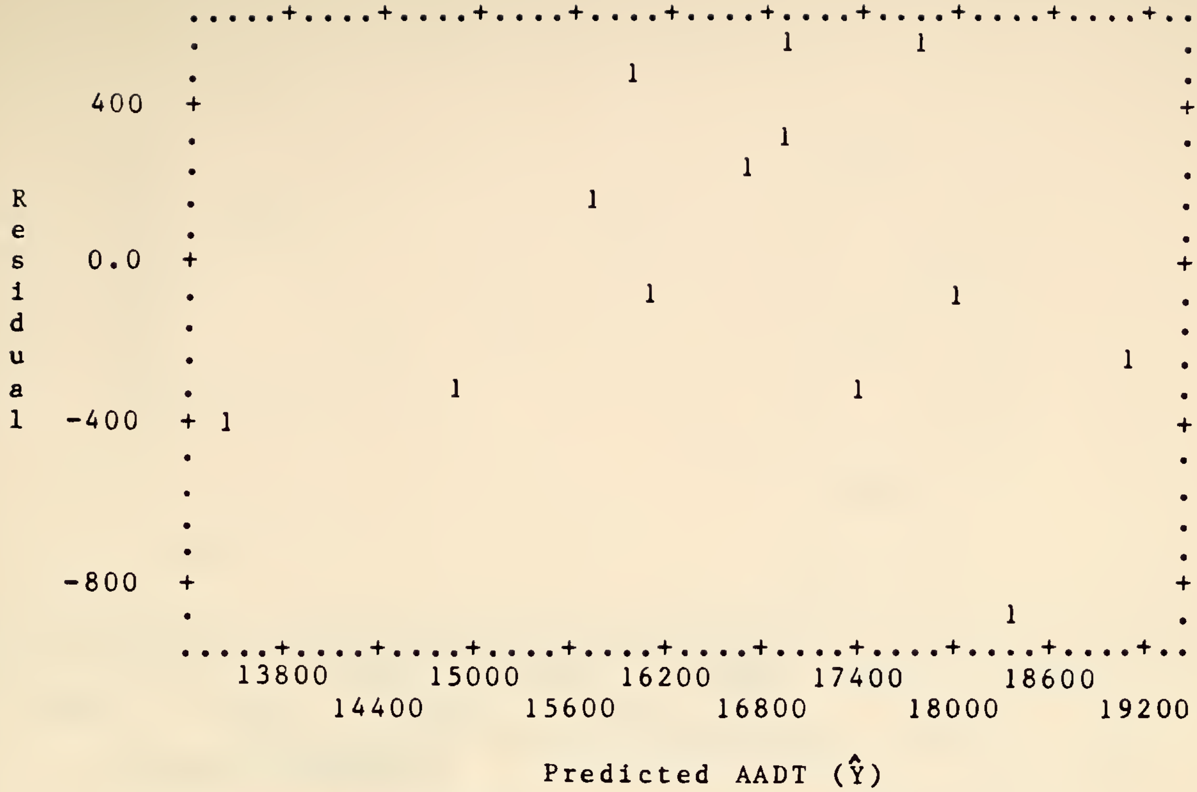


Figure E1.1: Residual Plot against $\hat{Y}$
(Station 3070A)

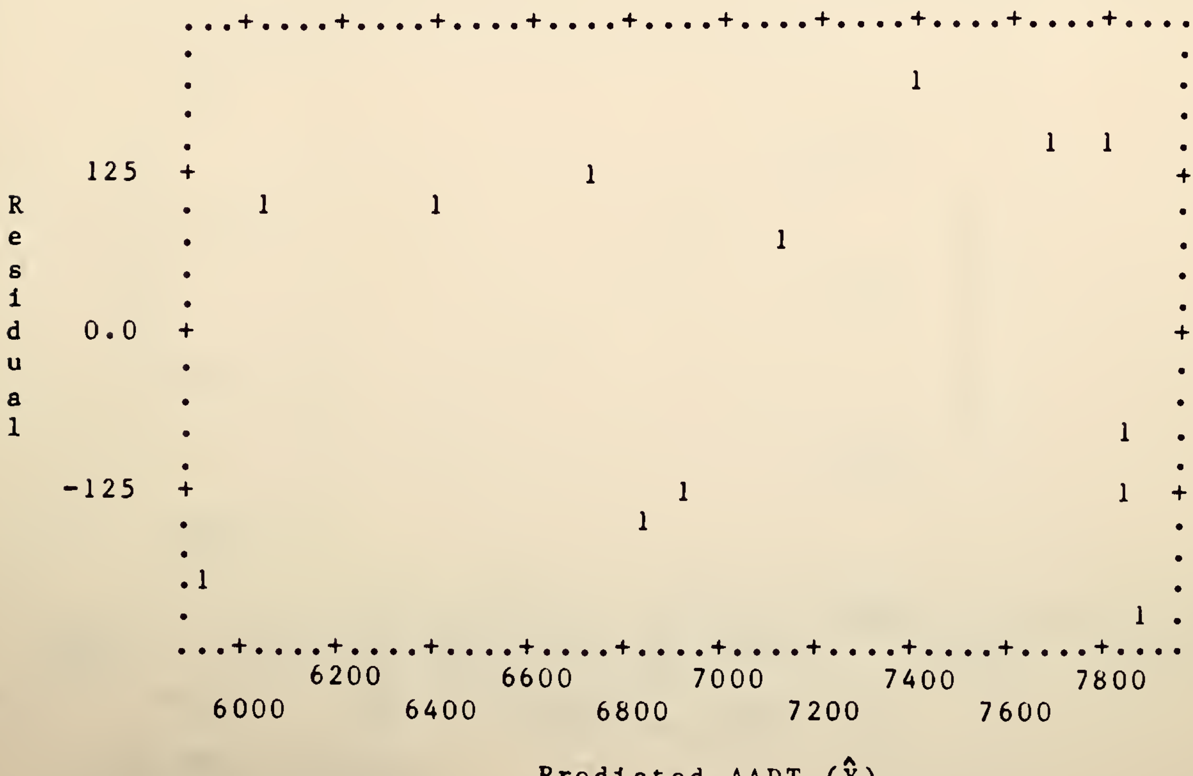


Figure E1.2: Residual Plot against $\hat{Y}$
(Station 68A)

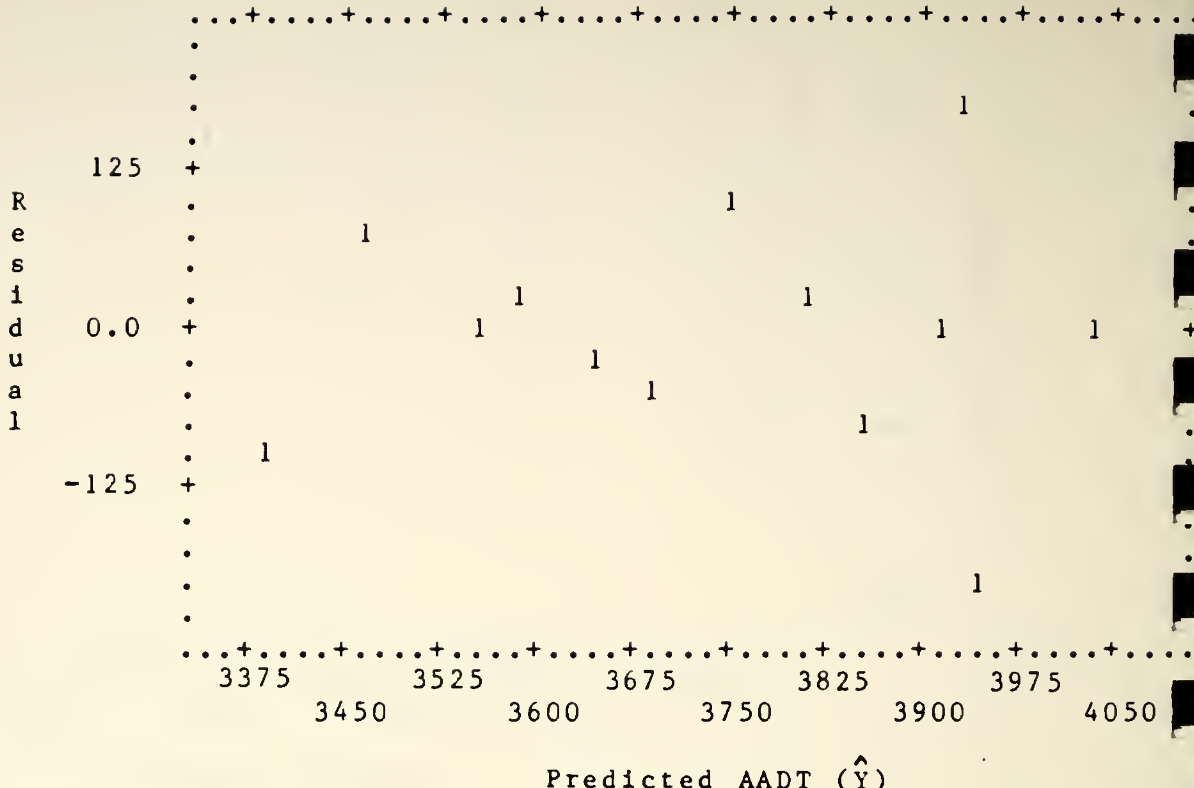


Figure E1.3: Residual Plot against $\hat{Y}$
(Station 301A)

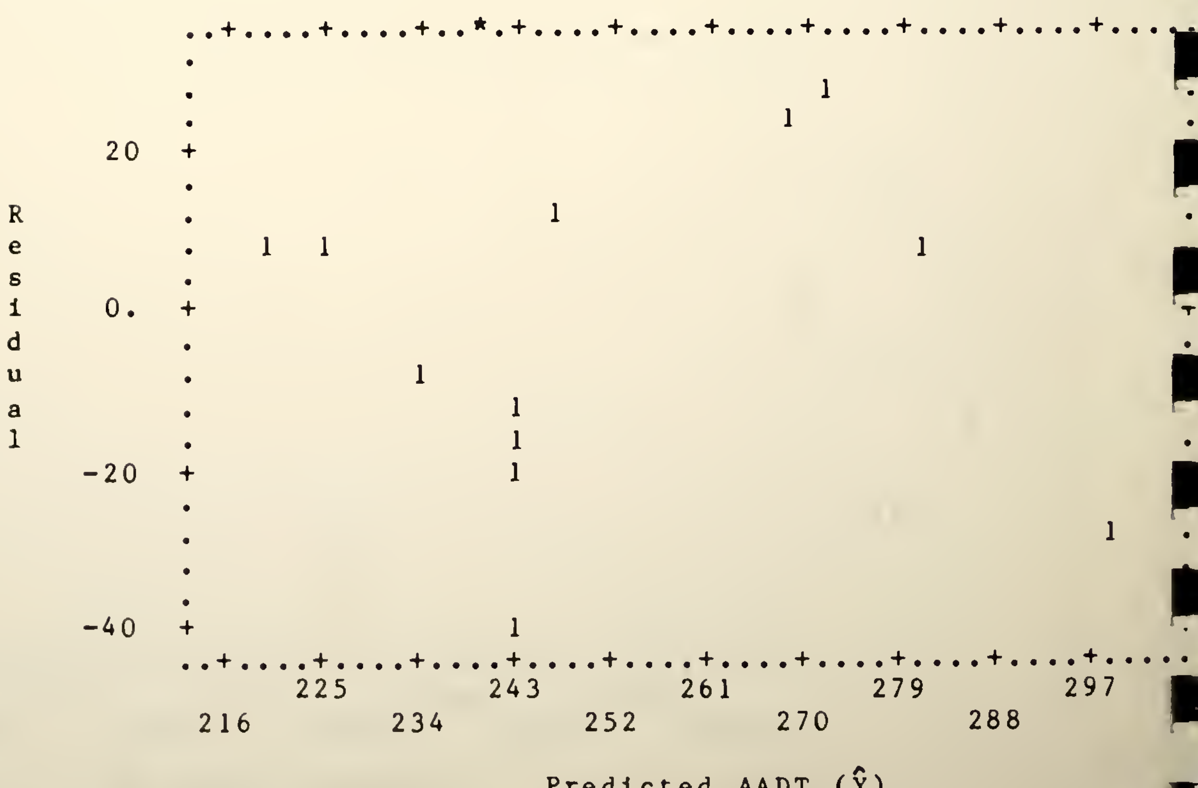


Figure E1.4: Residual Plot against $\hat{Y}$
(Station 7047A)

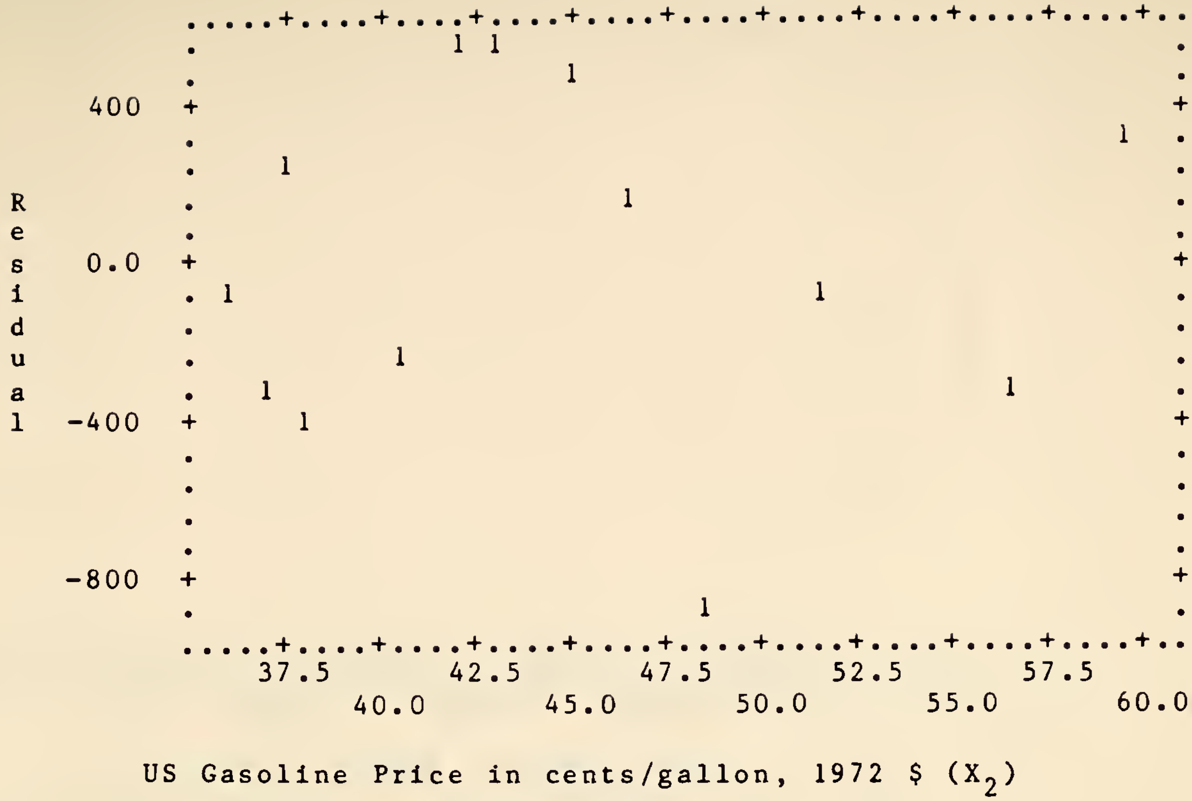


Figure E2.1.1: Residual Plot against X_2
(Station 3070A)

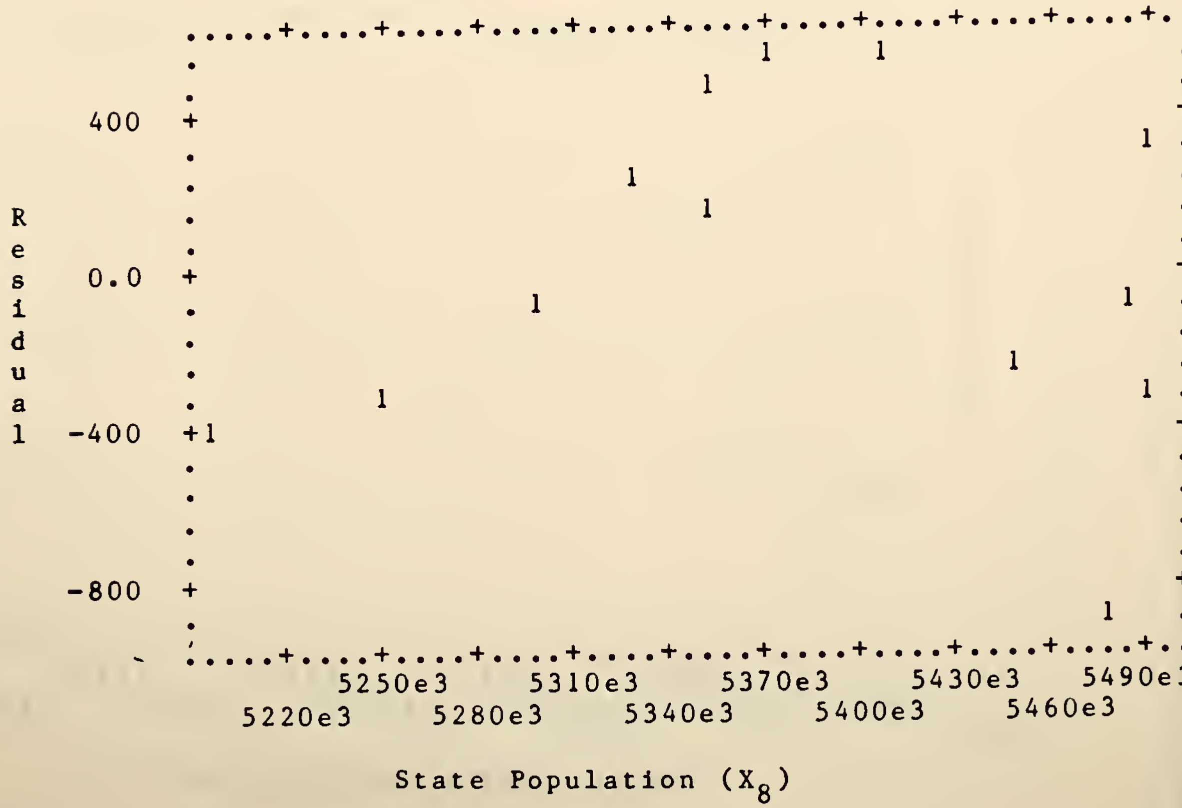


Figure E2.1.2: Residual Plot against X_8
(Station 3070A)

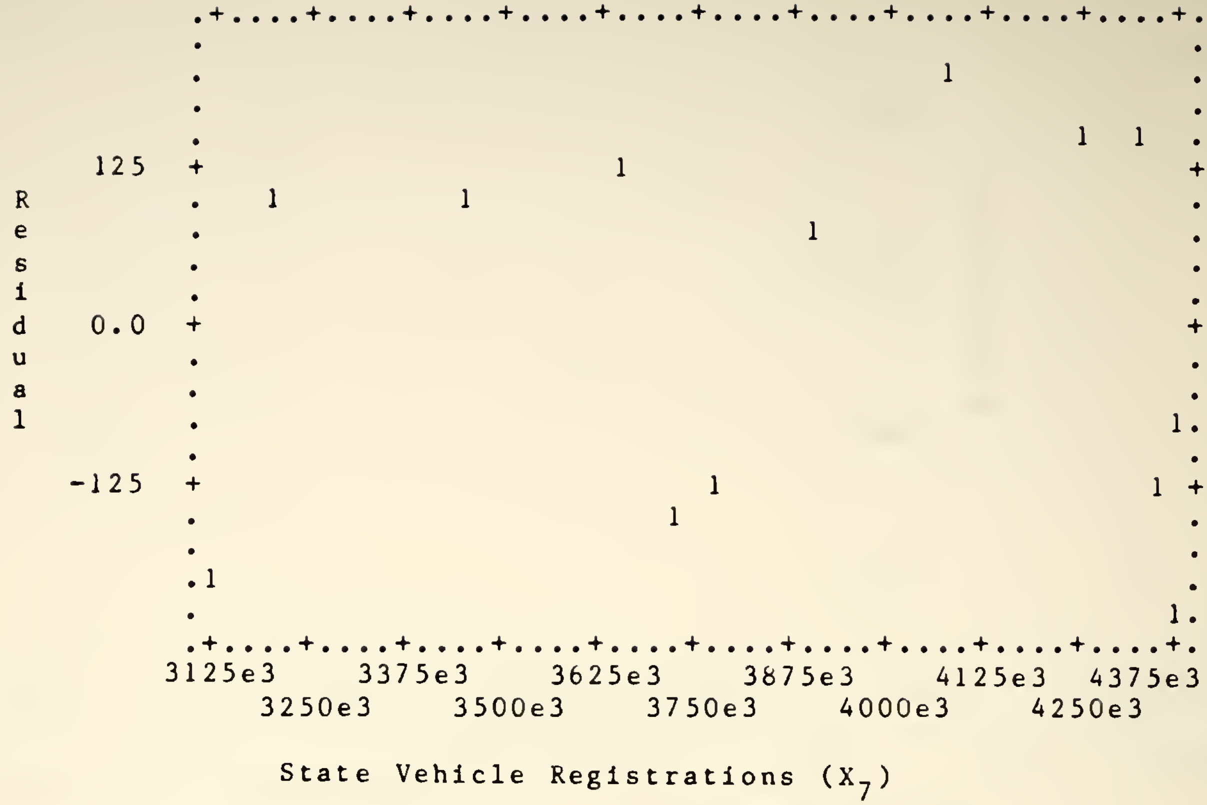


Figure E2.2.1: Residual Plot against X_7
(Station 68A)

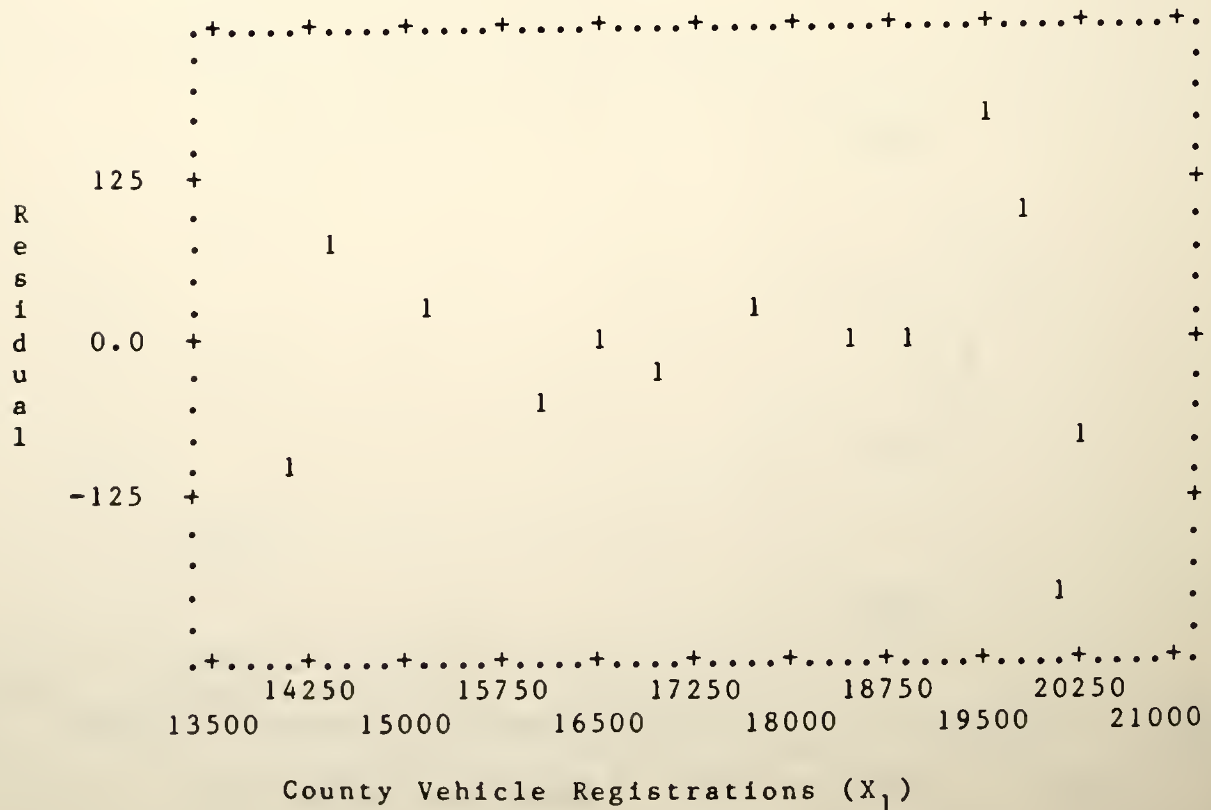


Figure E2.3.1: Residual Plot against X_1
(Station 301A)

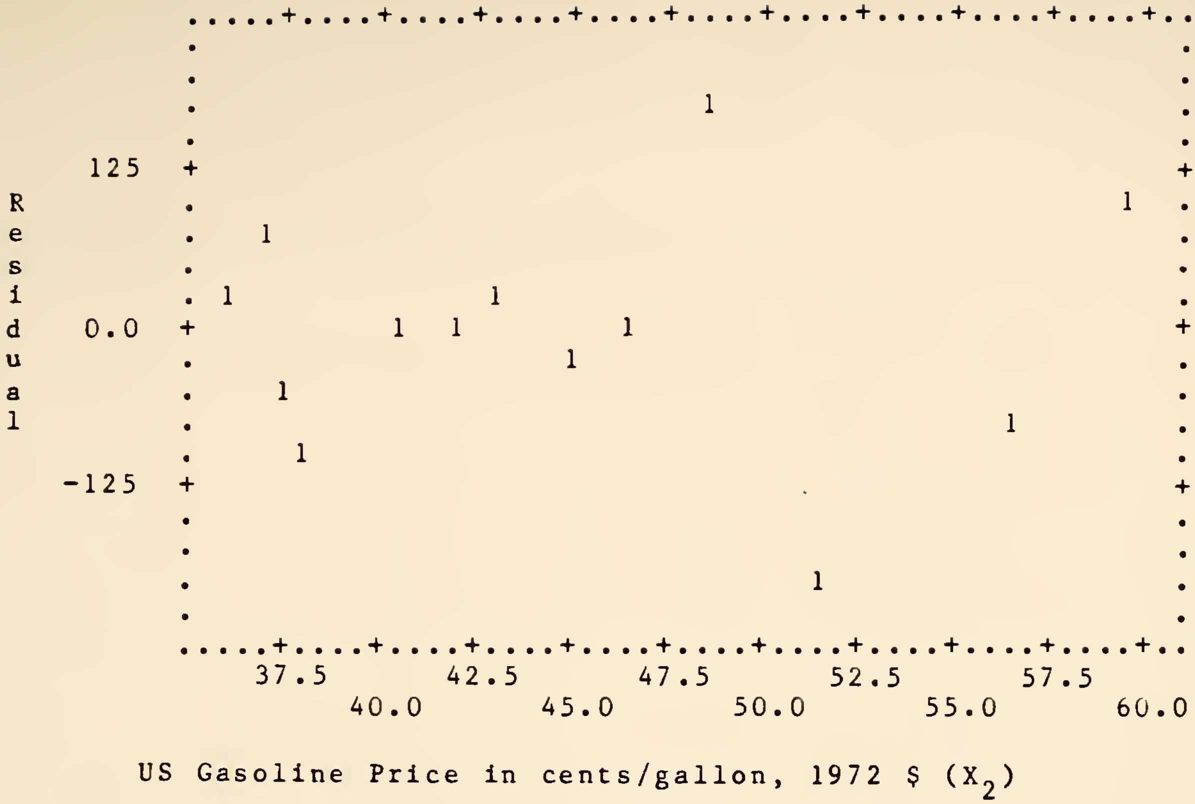


Figure E2.3.2: Residual Plot against X_2 (Station 301A)

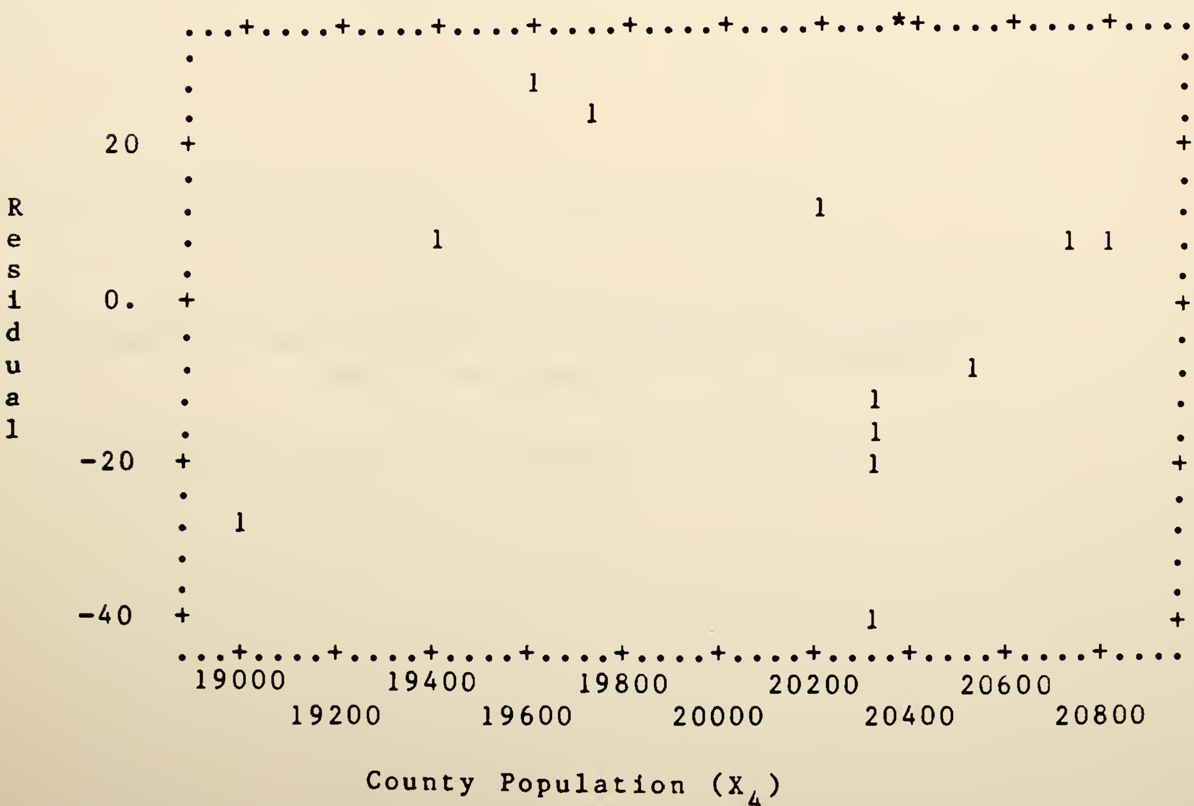


Figure E2.4.1: Residual Plot against X_4 (Station 7047A)

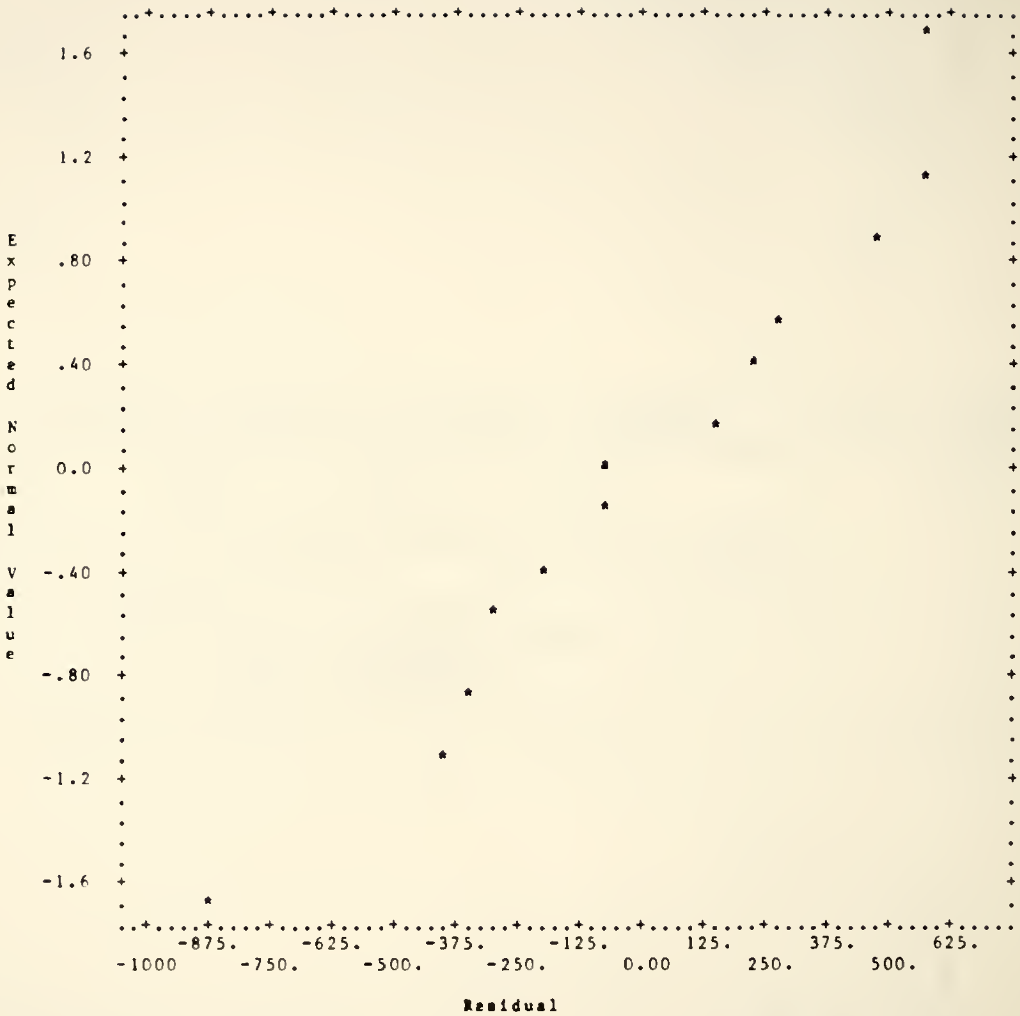


Figure E3.1: Normal Probability Plot of Residuals
(Station 3070A)

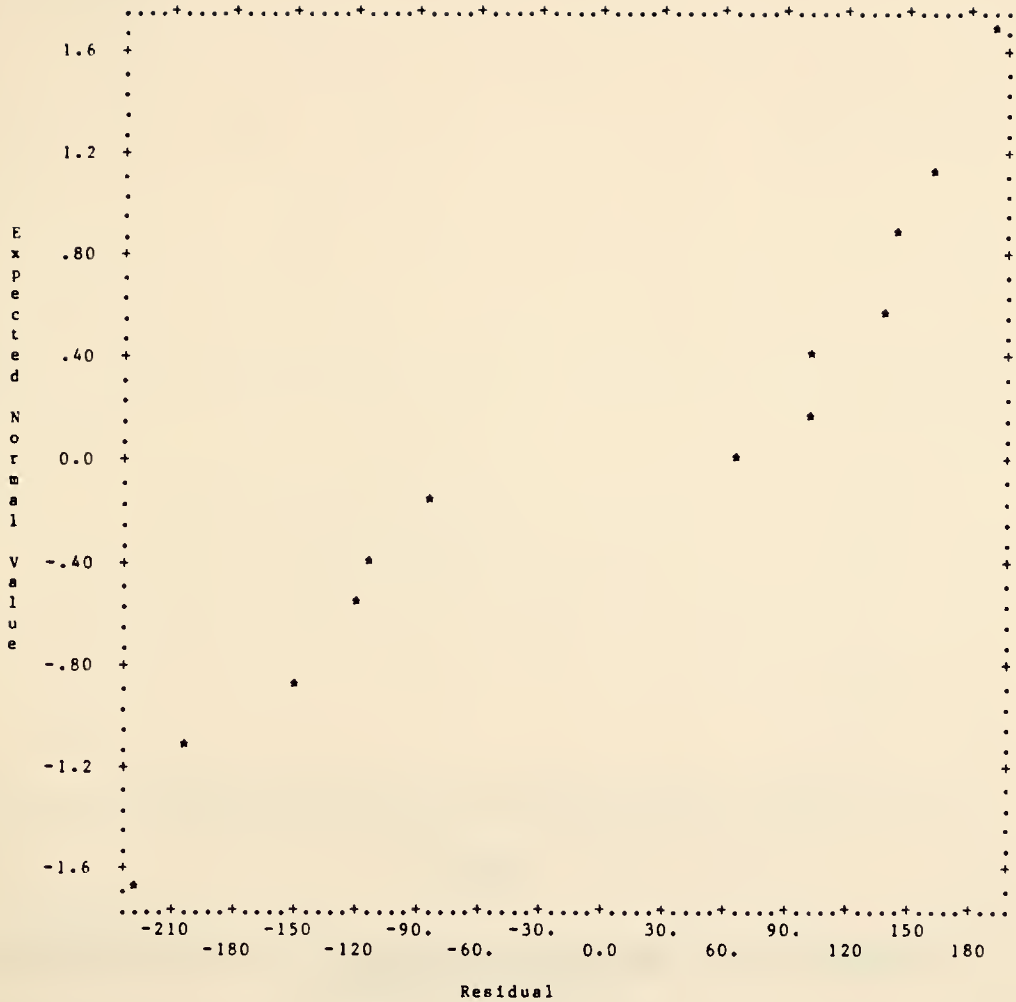


Figure E3.2: Normal Probability Plot of Residuals (Station 68A)

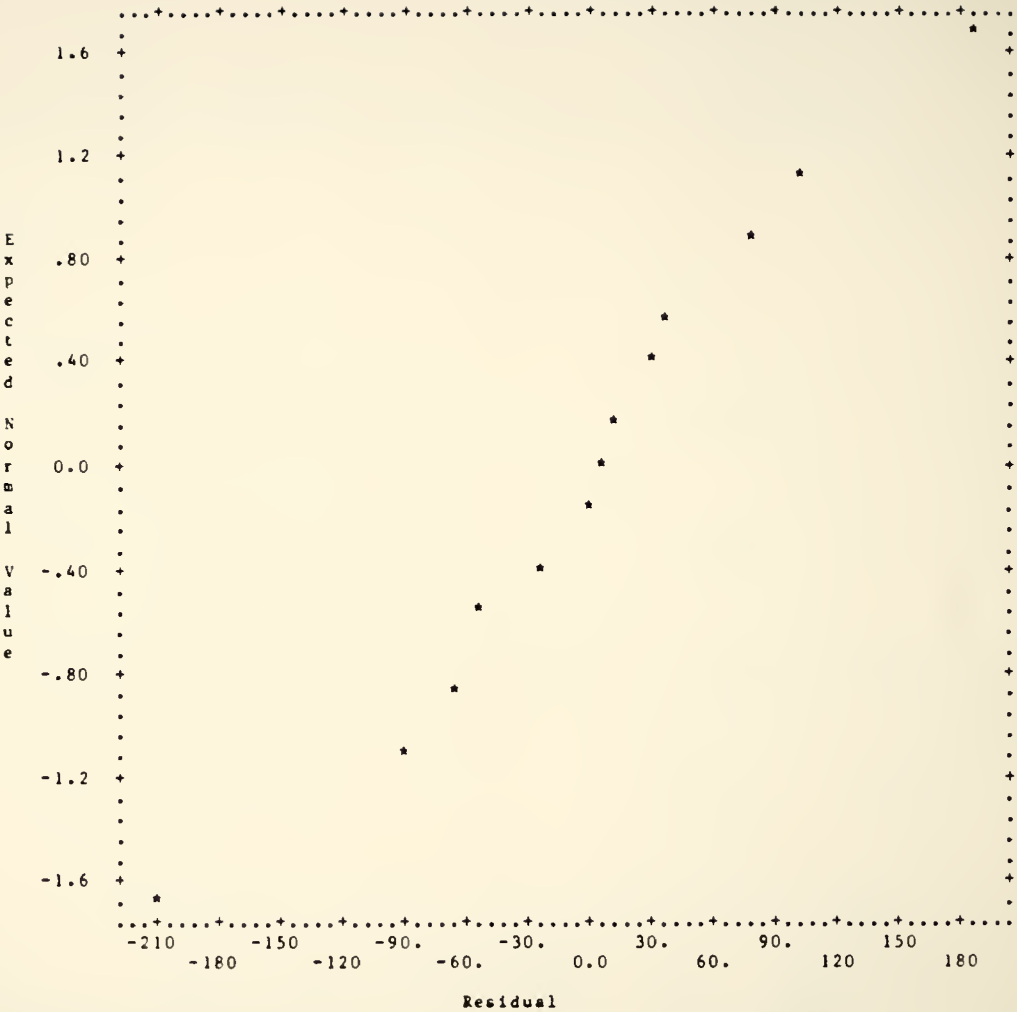


Figure E3.3: Normal Probability Plot of Residuals
(Station 301A)

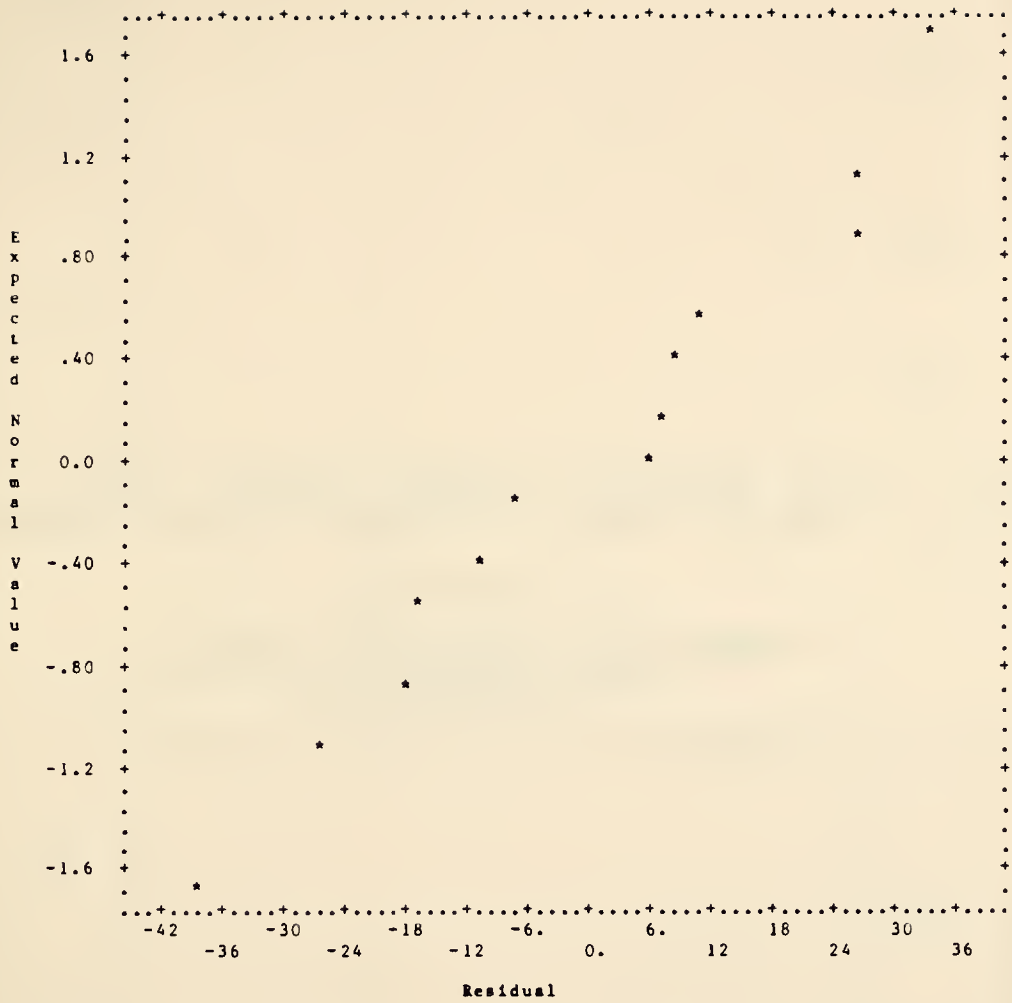


Figure E3.4: Normal Probability Plot of Residuals
(Station 7047A)

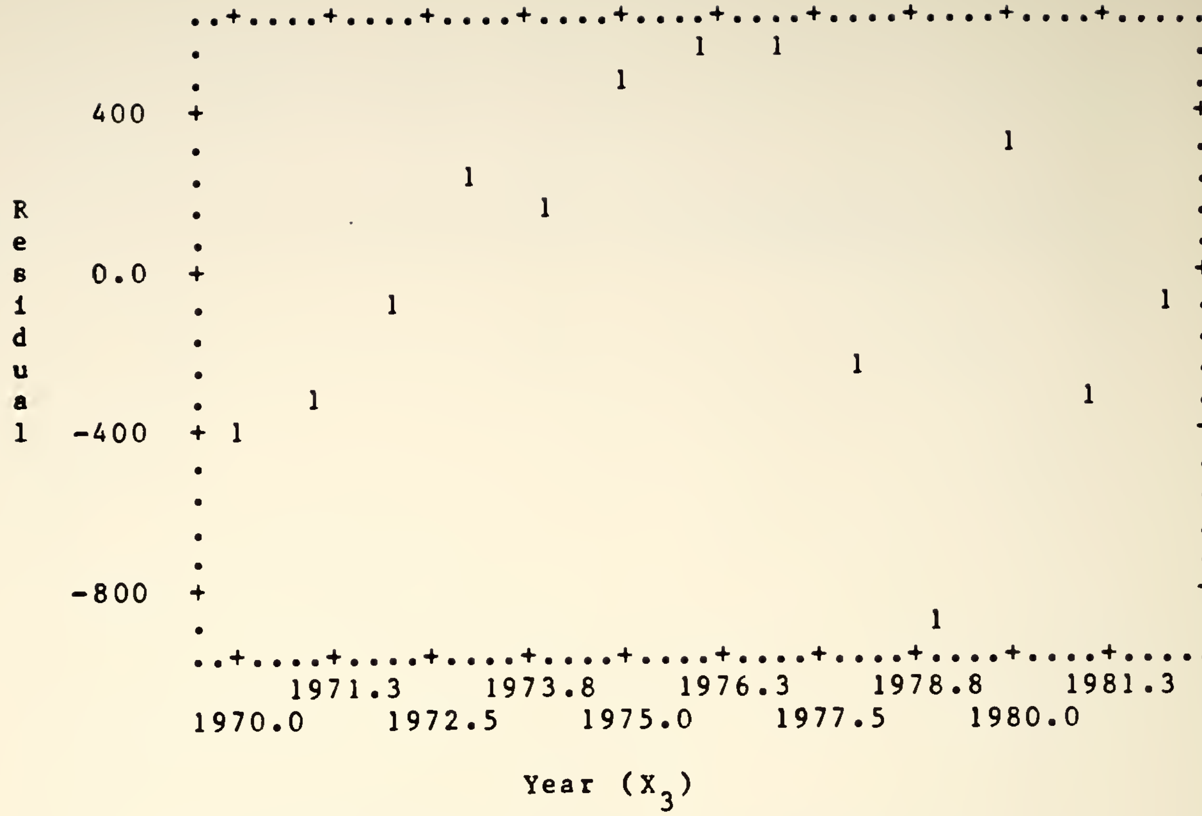


Figure E4.1: Residual Plot against X_3
(Station 3070A)

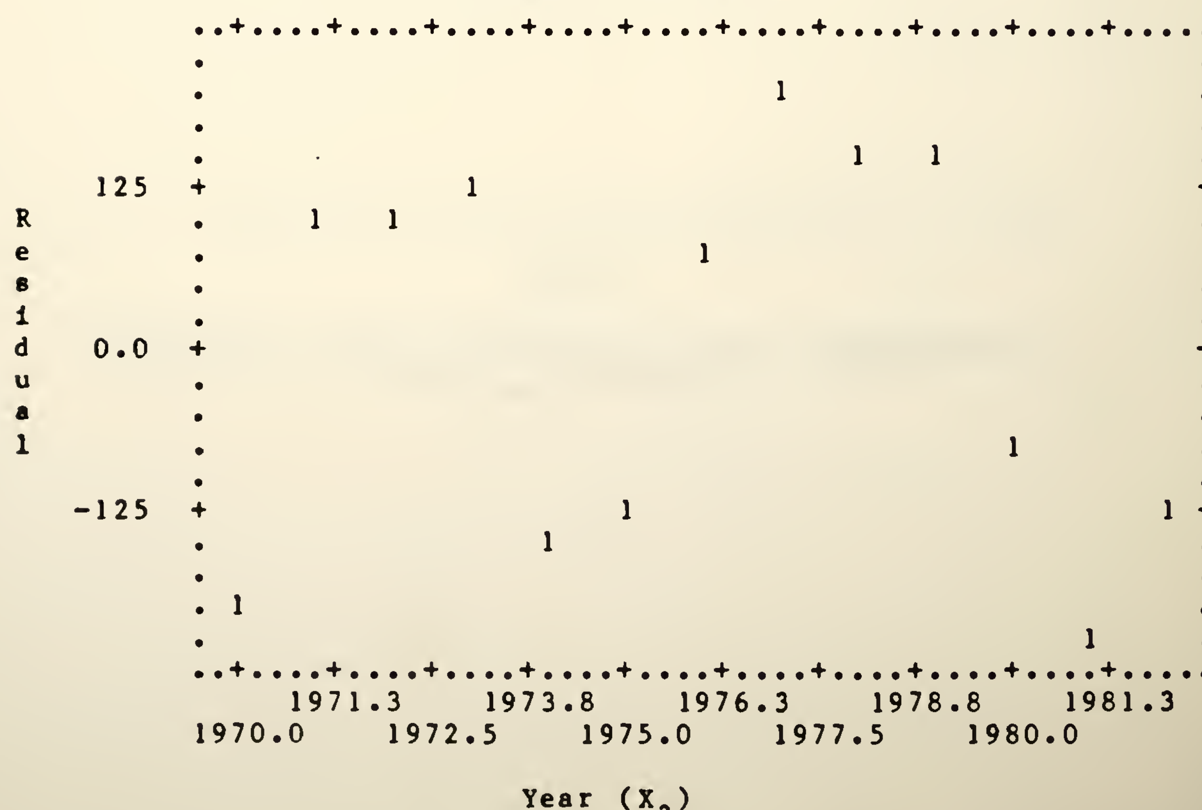


Figure E4.2: Residual Plot against X_3
(Station 68A)

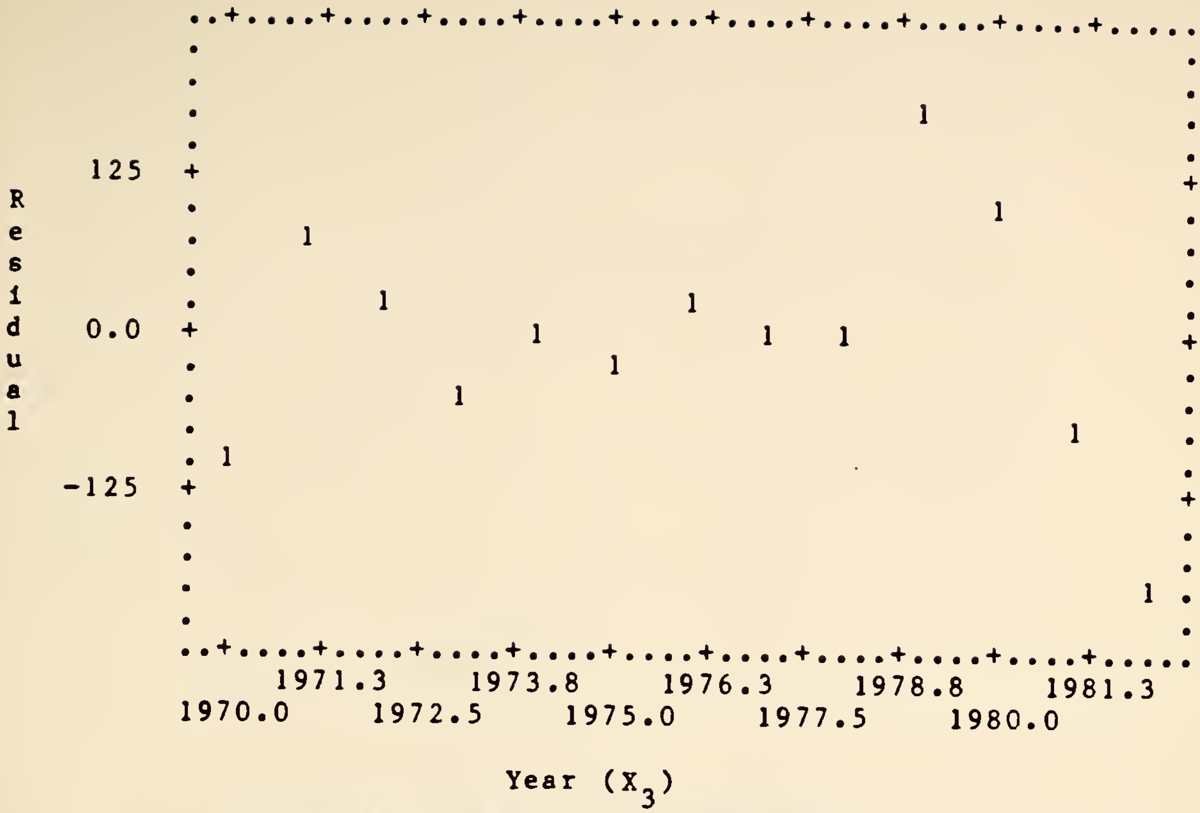


Figure E4.3: Residual Plot against X_3
(Station 301A)

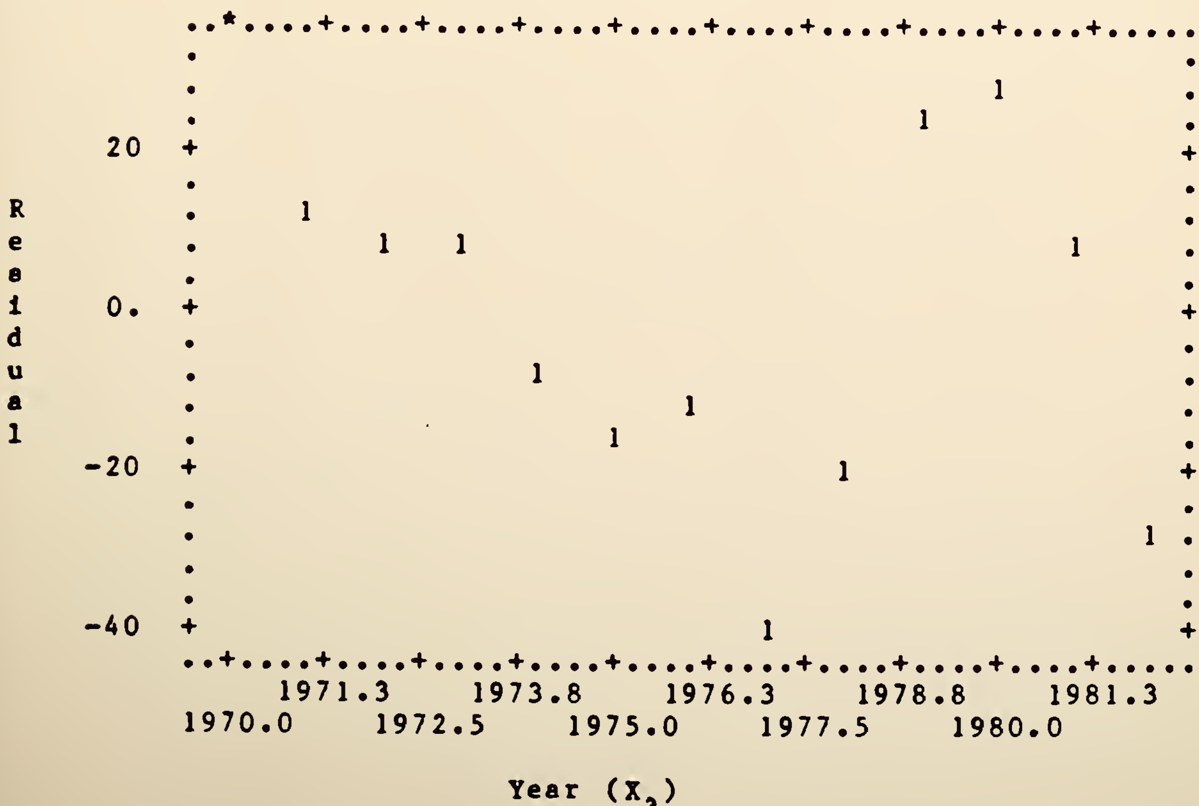


Figure E4.4: Residual Plot against X_3
(Station 7047A)

Appendix F

Examples on Simple Extrapolation

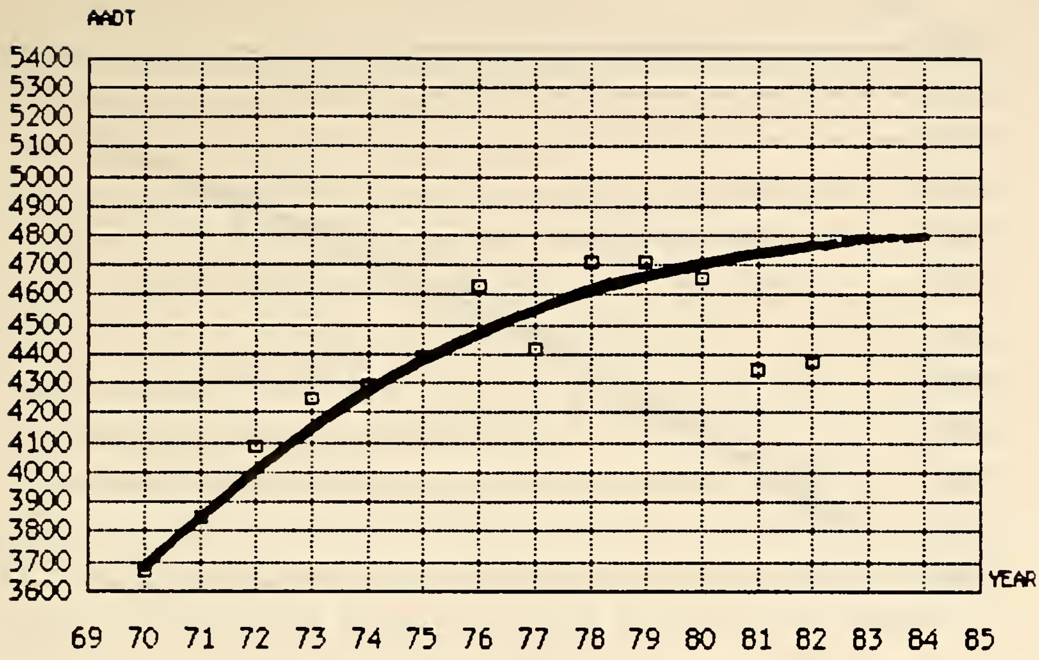


Figure F1: Simple Extrapolation of AADT
(Station 59A)

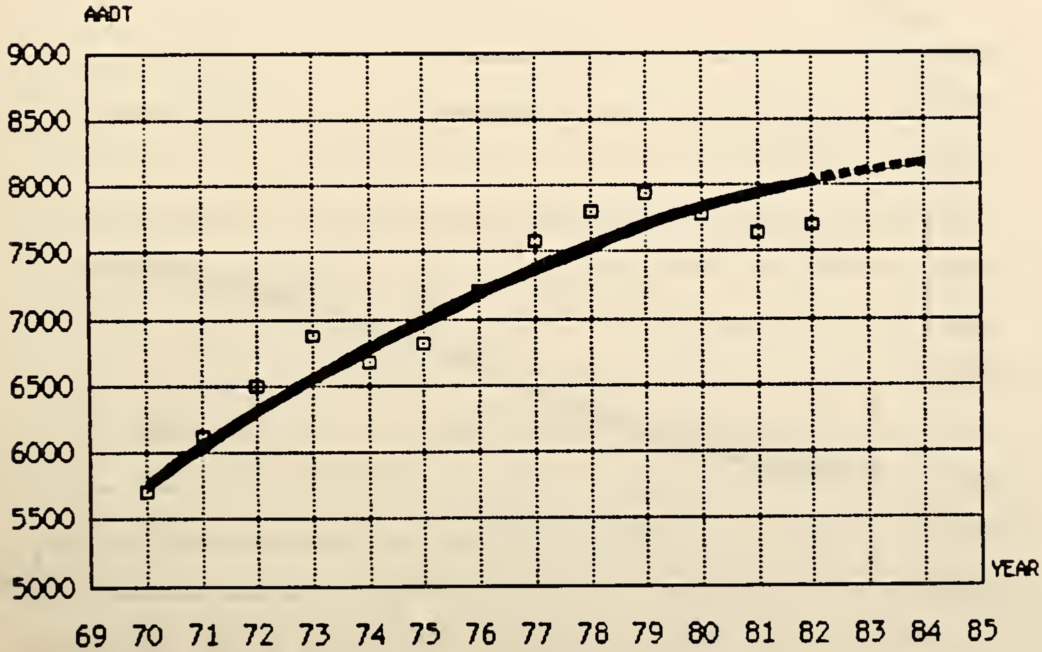


Figure F2: Simple Extrapolation of AADT
(Station 68A)

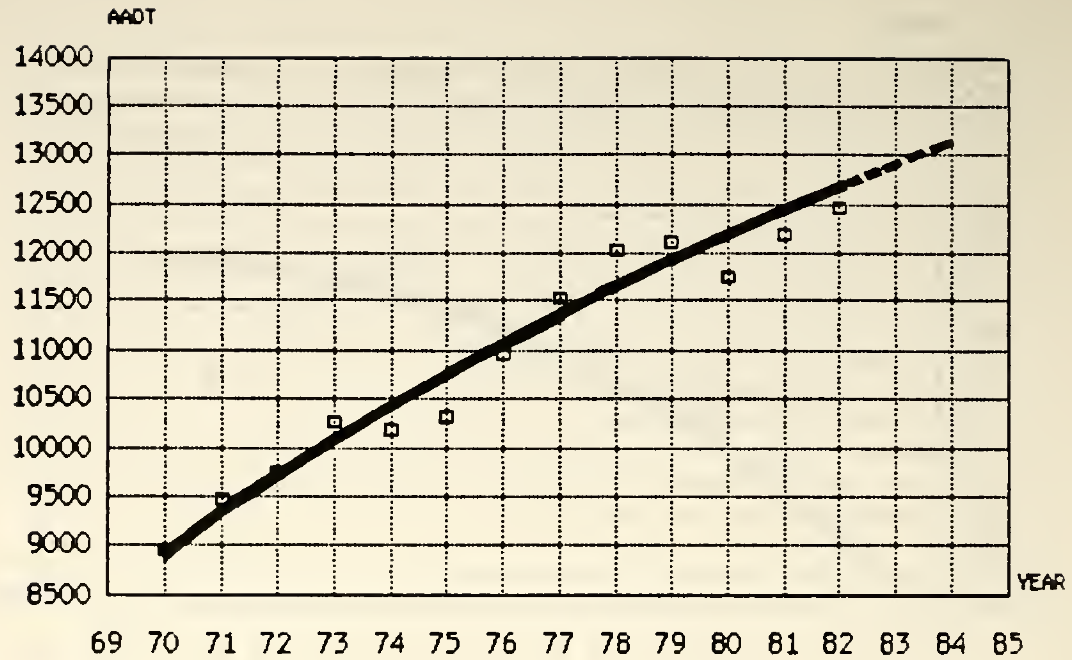


Figure F3: Simple Extrapolation of AADT
(Station 173A)

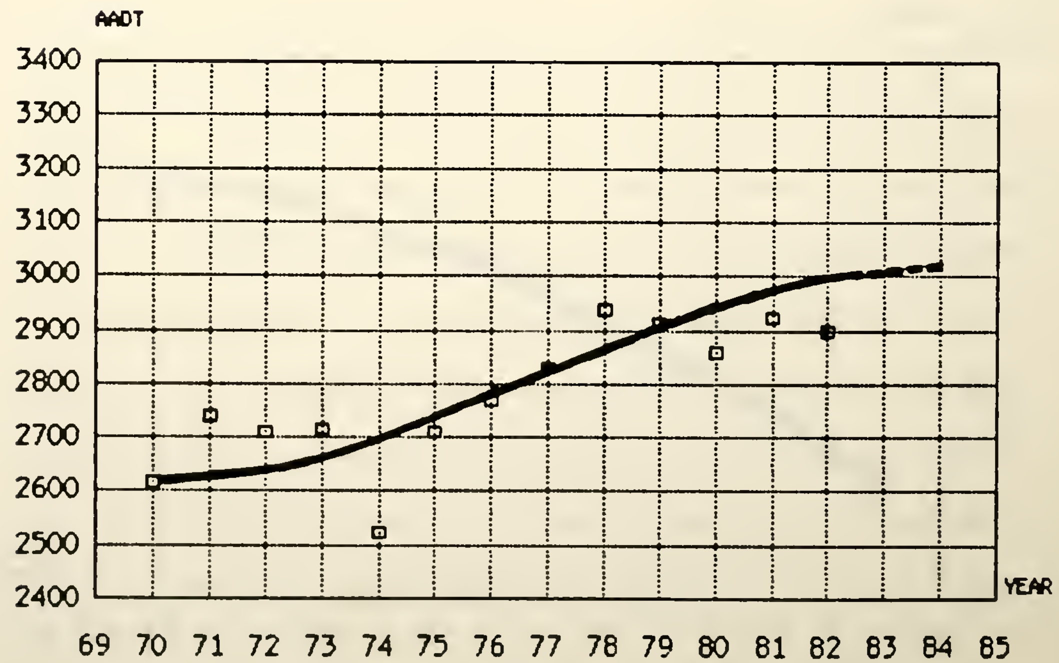


Figure F4: Simple Extrapolation of AADT
(Station 256A)

Appendix G

Statistical Test for Equality of Two Population Means

Test for Equality of Two Population Means [37]

Hypotheses

$$H_0 : \mu_1 = \mu_2$$

$$H_a : \mu_1 \neq \mu_2$$

where, μ_1 and μ_2 are two normal population means. Here, H_0 asserts that the two population means are the same, while H_a presumes they are not the same.

Evaluation of test statistic

Let $\bar{X}_1$ and $\bar{X}_2$ be the sample means of two independent samples. Estimators of the two population means are the sample means, which are calculated as follows:

$$\bar{X}_1 = \frac{\sum_{i=1}^{n_1} X_{1_i}}{n_1} \quad \text{and} \quad \bar{X}_2 = \frac{\sum_{i=1}^{n_2} X_{2_i}}{n_2}$$

where n_1 and n_2 are the number of samples for the two samples. The estimator of $\mu_1 - \mu_2$ is $\bar{X}_1 - \bar{X}_2$. An estimator of the common variance σ^2, denoted by s^2, is:

$$s^2 = \frac{\Sigma(X_{1_i} - \bar{X}_1)^2 + \Sigma(X_{2_i} - \bar{X}_2)^2}{n_1 + n_2 - 2}$$

An estimator of $\sigma^2(\bar{X}_1 - \bar{X}_2)$, denoted by $s^2(\bar{X}_1 - \bar{X}_2)$: the variance of sampling distribution of $\bar{X}_1 - \bar{X}_2$, is:

$$s^2(\overline{X}_1 - \overline{X}_2) = s^2\left|\frac{1}{n_1} + \frac{1}{n_2}\right|$$

Then, test statistic $t^* = \dfrac{\overline{X}_1 - \overline{X}_2}{s(\overline{X}_1 - \overline{X}_2)}$.

Decision Rule

Let t-value = $t(1 - \frac{\alpha}{2}; n_1 + n_2 - 2)$. Now, if $\left|t^*\right|$ < t-value, conclude H_0, i.e., two population means are same. Otherwise conclude H_a, i.e., two population means are not same. Here α is the level of significance (or degree of uncertainty). A value of 5 percent could be recommended for α. The term "$n_1 + n_2 - 2$" is known as degrees of freedom, where 2 degrees of freedom were lost to estimate two sample means.

Example

Two Rural Principal Arterial stations (68A and 254B) are used to demonstrate the principles described above. The data for this example are taken from Table A2 in Appendix A. Let the data of stations 68A and 254B represent samples of populations, indicated by the subscripts 1 and 2 in the discussion above. The values of the pertinent statistics and the decisions for the response variable AADT and the county level predictor variables are shown in Table G1. If the population means of the response variable (AADT in this case) and of the

Table G1

Tests for Equality of Variables Means for Two Locations

Variable	Key Statistics	Conclusion
AADT	$n_1 = 13$, $n_2 = 13$ $\bar{x}_1 = 7104$, $\bar{x}_2 = 7533$ $s^2 = 520147$ $\|t^*\| = 1.517$ t-value = 2.064	AADT of two stations are same
County Vehicle Registrations	$n_1 = 13$, $n_2 = 13$ $\bar{x}_1 = 24202$, $\bar{x}_2 = 30383$ $s^2 = 11917193$ $\|t^*\| = 4.565$ t-value = 2.064	County Vehicle Registrations of two counties are not same
County Population	$n_1 = 13$, $n_2 = 13$ $\bar{x}_1 = 31940$, $\bar{x}_2 = 37795$ $s^2 = 2951302$ $\|t^*\| = 8.689$ t-value = 2.064	County Population of two counties are not same
County Households	$n_1 = 13$, $n_2 = 13$ $\bar{x}_1 = 10350$, $\bar{x}_2 = 12662$ $s^2 = 961134$ $\|t^*\| = 6.012$ t-value = 2.064	County Households of two counties are not same
County Employment	$n_1 = 13$, $n_2 = 13$ $\bar{x}_1 = 7621$, $\bar{x}_2 = 10314$ $s^2 = 2927404$ $\|t^*\| = 4.013$ t-value = 2.064	County Employment of two counties are not same

county level predictor variables (employed in the proposed disaggregate model at one location) for the two locations are statistically the same, then the locations are "similar" and the disaggregate model is applicable at both locations. In Table G1, however, none of the predictor variables are statistically the same for the two stations. Thus, the stations are not "similar" and the disaggregate model developed for one station is not applicable at the other station.